Anwendungen des Empirischen Likelihood-Schätzers der Fehlerverteilung in AR(1)-Prozessen

Dissertation

zur Erlangung des akademischen Grades
„Doktor der Naturwissenschaften“

am Fachbereich
Mathematik und Informatik, Physik, Geographie
der Justus-Liebig-Universität Gießen

vorgelegt von
Michael Genz

Gießen 2004

D - 26

Dekan: Prof. Dr. Volker Metag

1. Berichterstatter: Prof. Dr. Erich Häusler (Gießen)
2. Berichterstatter: Prof. Dr. Winfried Stute (Gießen)

Datum der Disputation: 3. Dezember 2004

Inhaltsverzeichnis

Kapitel 1

Einleitung

In den Arbeiten von Owen (1988, 1990) wurde eine Idee in die mathematische Statistik eingeführt, die heute mit dem Schlagwort „empirical likelihood“ bezeichnet wird. Die Monografie von Owen (2001) faßt einen Teil der bisherigen Entwicklungen zu dieser Idee zusammen, die auch schon Eingang in die Lehrbuchliteratur gefunden hat; vgl. van der Vaart (1998), § 25.10 und Shao (1999), § 5.1.2. Das empirische Likelihood-Prinzip (kurz EL-Prinzip) ermöglicht im Fall von unabhängig und identisch verteilten Daten – unter anderem – bei Vorliegen „nichtparametrischer“ Informationen über eine unbekannte Verteilung, wie z.B. der Kenntnis des Erwartungswerts, des Medians oder der Varianz, die Konstruktion von nichtparametrischen Verteilungsschätzern, die im Gegensatz zur klassischen empirischen Verteilungsfunktion F_n die gegebene Information ausnutzen, indem die Wahrscheinlichkeitsgewichte $1/n$ in F_n durch zufallsabhängige Gewichte ersetzt werden. Ist etwa bekannt, daß eine ansonsten unbekannte Verteilung zentriert ist, so heißt dies, daß auch der Verteilungsschätzer stets zentriert ist. Mit der Anwendung des EL-Prinzips in genau dieser Situation beschäftigt sich die vorliegende Arbeit, und zwar nicht nur im Rahmen von unabhängig und identisch verteilten Zufallsvariablen. Auf Grund der besseren Anpassung des EL-Schätzers an das Modell, bedingt durch die Einbeziehung der Zusatzinformation in die Schätzung, sind im allgemeinen bessere Ergebnisse in statistischen Anwendungen als bei Verwendung von F_n zu erwarten.

Qin und Lawless (1994) sowie Zhang (1997) betrachten die auf dem EL-Prinzip beruhende Schätzung der unbekannten Verteilungsfunktion F einer Folge von unabhängig und identisch verteilten Zufallsvariablen, wenn Informationen der funktionalen Form $\gamma(\mathrm{F}) = 0$ vorliegen. Dabei lassen sie in γ sogar noch einen zusätzlichen Parameter θ zu. Diese Möglichkeit spielt für uns aber keine Rolle. Wie oben bereits erwähnt interessiert uns nämlich nur die Zusatzinformation der Zentriertheit von F, also das Funktional

$$\gamma(\mathrm{F}) = \int_{\mathbb{R}} x\mathrm{F}(dx).$$

Die zugehörige, nach dem EL-Prinzip modifizierte empirische Verteilungsfunktion sei mit $\mathrm{F}_{n,z}$ bezeichnet. Das Hauptergebnis in Zhang (1997) enthält für den empirischen Prozeß zu $\mathrm{F}_{n,z}$ einen funktionalen Grenzwertsatz mit einem zentrierten Gaußprozeß als Grenzprozeß. Dessen Varianzfunktion ist punktweise nichtgrößer als die der mit F zeittransformierten Brown'schen Brücke, also des Grenzprozesses im funktionalen Grenzwertsatz von Donsker für den klassischen empirischen Prozeß zu F_n. In Verbindung mit Beispiel 2 im Abschnitt 5.3 von Bickel et al. (1998) impliziert dieses Resultat die asymptotische Effizienz des EL-Verteilungsschätzers $\mathrm{F}_{n,z}$ im Sinne des Hajek'schen Faltungstheorems. Damit liefert das EL-Prinzip im Fall von unabhängig und identisch verteilten Daten einen in diesem Sinne asymptotisch optimalen Verteilungsschätzer.

Eine in der Literatur bisher noch nicht untersuchte Anwendung des EL-Prinzips ist seine Verwendung bei der Konstruktion von Anpassungstests auf parametrische Verteilungsfamilien. Durbin (1973) zeigte im Falle unabhängig und identisch verteilter Zufallsvariablen einen für solche Tests grundlegenden funktionalen Grenzwertsatz für den klassischen empirischen Prozeß mit geschätztem Parameter. Im Kapitel 2 der vorliegenden Arbeit wird – unter der Zusatzbedingung der Zentriertheit der Zufallsvariablen – ein analoger Grenzwertsatz für den empirischen Prozeß mit geschätztem Parameter zu $\mathrm{F}_{n,z}$ bewiesen. Dabei zeigt sich eine Verringerung der Varianz des Grenzprozesses gegenüber dem Durbin'schen Resultat, wenn der Verteilungsparameter mit dem Maximum-Likelihood-Verfahren geschätzt wird.

Während die Annahme der Zentriertheit unabhängig und identisch verteilter Daten eher wenig anwendungsrelevant erscheinen mag, ist die Zentriertheit der Fehlervariablen ein wesentlicher Bestandteil vieler Zeitreihenmodelle. Da hier die Fehler im allgemeinen nicht beobachtbar sind, kann ihre unbekannte Verteilung F nicht mittels der empirischen Verteilungsfunktion F_n der Fehler geschätzt werden. Deshalb wird in der Literatur meist die empirische Verteilungsfunktion $\mathrm{F}_{n,res}$ auf Grund von Residuen als Schätzer für F verwendet. Wegen der Zentriertheit der Fehler ist es jedoch naheliegend, anstelle von $\mathrm{F}_{n,res}$ eine nach dem EL-Prinzip modifizierte Form $\mathrm{F}_{n,res,z}$ zu verwenden.

Wichtige Zeitreihenmodelle sind z.B. autoregressive Prozesse (AR(p)-Prozesse der Ordnung $p \geq 1$), in denen die Fehlervariablen üblicherweise als unabhängig und identisch verteilt sowie zentriert vorausgesetzt werden. Der Größe des Autoregressionsparameters ρ fällt dabei bekanntermaßen eine entscheidende Rolle für das stochastische Verhalten der Prozesse zu. Im Fall von AR(1)-Prozessen (mit $\rho \in \mathbb{R}$) hat man den *stabilen* (oder *stationären*) Fall $|\rho| < 1$, den *semistabilen* Fall $|\rho| = 1$ und den *explosiven* Fall $|\rho| > 1$ zu unterscheiden. Bei beliebiger Ordnung p (mit $\rho \in \mathbb{R}^p$) erfolgt die entsprechende Klassifikation über die Beträge der im allgmeinen komplexwertigen Nullstellen des charakteristischen Polynoms des Prozesses, und es sind Mischformen möglich, die eine vollständige Analyse sehr komplex gestalten. Wir werden uns daher in der vorliegenden Arbeit auf das Studium von AR(1)-Prozessen beschränken und zwei Anwedungen des EL-Prinzips im Zusammenhang mit der Schätzung der Fehlerverteilung untersuchen.

Für AR(p)-Prozesse mit $p \geq 1$ haben Boldin (1982) und später Koul (1992) im stabilen Fall für $\mathrm{F}_{n,res}$ und F_n die asymptotische Äquivalenz bewiesen. Entsprechende Ergebnisse im semistabilen und im explosiven Fall findet man bei Ling (1998) beziehungsweise Koul und Levental (1989), wobei sich die letzte Arbeit auf den Fall $p = 1$ beschränkt. Lee (1997) ist unseres Wissen die einzige Arbeit, die den auf $\mathrm{F}_{n,res}$ basierenden empirischen Prozeß mit geschätztem Parameter behandelt. Die Autorin studiert autoregressive Prozessen beliebiger Ordnung $p \geq 1$, beschränkt sich allerdings auf den Fall strikt stationärer Prozesse und engt das Modell stark ein, indem sie als Verteilungsfamilien nur Skalenfamilien zuläßt. In diesem Fall zeigt sie einen funktionalen Grenzwertsatz, der dem von Durbin (1973) entspricht. Wir betrachten in Kapitel 3 zwar nur AR(1)-Prozesse, werden dafür aber beliebige Verteilungsparameter zulassen und einen von der Größe von ρ unabhängigen Grenzwertsatz für den empirischen Prozeß mit geschätztem Parameter zu $\mathrm{F}_{n,res}$ beweisen, der ein vollwertiges Analogon zu dem in Durbin (1973) bewiesenen Ergebnis im AR(1)-Modell ist. Über den zur Konstruktion von $\mathrm{F}_{n,res}$ benötigten Schätzer für ρ werden dabei nur Annahmen über die Geschwindigkeit seiner Konsistenz gemacht, die z.B. der Kleinst-Quadrate-Schätzer (mit einer geeigneten Modifikation für $|\rho| = 1$) erfüllt.

Wie bereits oben erläutert bietet sich die Verwendung des EL-Prinzips zur Modifikation von $\mathrm{F}_{n,res}$ zu $\mathrm{F}_{n,res,z}$ an. Deshalb werden wir in Kapitel 3 auch den empirischen Prozeß mit geschätztem Parameter basierend auf $\mathrm{F}_{n,res,z}$ studieren. Dabei zeigen wir Verteilungskonvergenz dieses Prozesses gegen den gleichen Grenzprozeß, den wir schon im Fall unabhängig und identisch verteilter zentrierter Variablen erhalten haben. Die entsprechenden Ergebnisse aus Kapitel 2 erweisen sich dabei als wichtige Beweishilfsmittel. Es ergibt sich damit sowohl für $\mathrm{F}_{n,res}$ wie für $\mathrm{F}_{n,res,z}$ ein für alle ρ gleiches asymptotisches Verteilungsverhalten, ganz im Gegenteil zu den Verhältnissen bei der Schätzung von ρ: Der Kleinst-Quadrate-Schätzer z.B. hat ja in den drei Fällen $|\rho| < 1$, $|\rho| = 1$ und $|\rho| > 1$ völlig verschiedene Grenzverteilungen.

Also konvergiert sowohl im Modell unabhängig und identisch verteilter zentrierter Zufallsvariablen als auch im AR(1)-Modell der durch das EL-Verfahren gewonnene empirische Prozeß mit geschätztem Parameter gegen den selben Grenzprozeß, nämlich einen zentrierten Gaußprozeß mit einer Varianzfunktion, die punktweise nichtgrößer ist als die des klassischen Grenzprozesses von Durbin (1973). Allerdings können die erzielten Ergebnisse nicht direkt zur Konstruktion kritischer Werte für Tests herangezogen werden, da die Grenzverteilung in nichttrivialer Weise von der unbekannten Verteilung F abhängt, so daß aus dem empirischen Prozeß abgeleitete Teststatistiken, z.B. vom Kolmogorov-Smirnov-Typ, im allgemeinen nicht einmal asymptotisch verteilungsfrei sind.

Eine Möglichkeit, dennoch kritische Werte zu erhalten, ist das Bootstrap-Verfahren; vgl. z.B. Shao und Tu (1995). Die Tatsache, daß F unter der Hypothese des Anpassungstests einer parametrischen Verteilungsfamilie entspringt, ermöglicht die Durchführung eines parametrischen Bootstrap-Ansatzes. Im Modell unabhängig und identisch verteilter Zufallsvariablen wurden Bootstrapversionen des Grenzwertsatzes aus Durbin (1973) bereits von Stute et al. (1993) sowie später von Babu und Rao (2004) bewiesen. Sowohl für den auf $\mathrm{F}_{n,z}$ beruhenden empirischen Prozeß mit geschätztem Parameter im Modell unabhängig und identisch verteilter Zufallsvariablen als auch für die aus $\mathrm{F}_{n,res}$ und $\mathrm{F}_{n,res,z}$ abgeleiteten empirischen Prozesse mit geschätztem Parameter im AR(1)-Modell beziehungsweise die zugehörigen funktionalen Grenzwertsätze sind entsprechende Bootstrap-Resultate aus der Literatur nicht bekannt. Da die hierfür benötigten Beweismethoden stets die selben sind, wird in Kapitel 4 der Beweis exemplarisch für den empirischen Prozeß mit geschätztem Parameter zu $\mathrm{F}_{n,res,z}$ im AR(1)-Modell ausführlich dargestellt.

In der Literatur gibt es bereits einige Anwendungen des EL-Prinzips im Zeitreihen-Kontext. Monti (1997) verwendet es zur Schätzung von Parametern im Zusammenhang von Spektraldichten im Fall stationärer Prozesse. Auch in der Arbeit von Chuang und Chan (2002) werden Schätzer und Tests im Zusammenhang mit den Koeffizienten von AR-Prozessen behandelt. Schließlich verwenden Müller, Schick und Wefelmeyer (2004) das EL-Prinzip zur Kerndichteschätzung.

Eine weitere Anwendungsmöglichkeit des EL-Prinzips in AR(1)-Modellen ist die Konstruktion von nichtparametrischen Prognoseintervallen bzw. Prognoseschranken. Im Fall $|\rho| > 1$ untersuchen Stute und Gründer (1993) nichtparametrische Prognoseintervalle für zukünftige Zeitreihenwerte auf Basis von Quantilen $\mathrm{F}_{n,res}$. Im Fall $|\rho| < 1$, der in der vorliegenden Arbeit betrachtet wird, zeigt sich im Vergleich zu diesen Ergebnissen allerdings ein wesentlich anderes asymptotisches Verhalten der Prognoseintervalle.

Im Kapitel 5 werden zunächst Ein-Schritt-Prognosen behandelt. Wir werden sehen, daß die mithilfe von $\mathrm{F}_{n,res,z}$ definierten Intervalle beziehungsweise die zugehörigen bedingten Überdeckungswahrscheinlichkeiten $\mathrm{P}_{n,res,z,\alpha}$ zu einem vorgegebenen Niveau α ebenso konsistent sind wie die auf $\mathrm{F}_{n,res}$ aufbauenden $\mathrm{P}_{n,res,\alpha}$. Auch die Länge der Intervalle ist asymptotisch gleich. Unterschiede zeigen sich jedoch beim Vergleich der Folgen

$$\sqrt{n}(\mathrm{P}_{n,res,\alpha} - (1-\alpha))$$

und

$$\sqrt{n}(\mathrm{P}_{n,res,z,\alpha} - (1-\alpha)).$$

Für beide stellt sich Verteilungskonvergenz heraus, jedoch hat die Grenzvariable im Falle der EL-Schätzung eine nichtgrößere Varianz. Kapitel 6 überträgt die obigen Ergebnisse auf Zwei-Schritt-Prognoseintervalle und -schranken, um die methodischen Unterschiede bei der Diskussion von Ein- und Mehr-Schritt-Prognosen aufzuzeigen.

Im letzten Kapitel der Arbeit schließlich werden die theoretischen (asymptotischen) Aussagen der Arbeit im Sinne einer adäquaten Approximation bei endlichen Stichprobenumfängen in konkreten Beispielen empirisch im Rahmen von Simulationsstudien überprüft. Außerdem werden die Kolmogorov-Smirnov-Tests zu den empirischen Prozessen mit geschätzten Parametern zu $\mathrm{F}_{n,z}$ beziehungsweise $\mathrm{F}_{n,res,z}$ einerseits und F_n beziehungsweise $\mathrm{F}_{n,res}$ andererseits bezüglich ihrer Güte verglichen. Als Konsequenz der in Kapitel 2 und 3 bewiesenen funktionalen Grenzwertsätze weisen dabei die EL-basierten Tests eine größere Güte auf als die klassischen.

Ich danke Herrn Prof. Dr. Erich Häusler für die Anregung zum Thema dieser Dissertation und die hervorragende Betreuung während ihrer Anfertigung.

Kapitel 2

Der empirische Prozeß mit geschätztem Parameter bei unabhängigen zentrierten Zufallsvariablen

2.1 Modellvoraussetzungen und bekannte Ergebnisse

Wir betrachten in diesem Kapitel eine Folge von unabhängig und identisch nach einer Verteilungsfunktion F verteilten Zufallsvariablen $(e_i)_{i\in\mathbb{N}}$. Die Bezeichnung der Zufallsvariablen ist auf den ersten Blick etwas ungewöhnlich. Sie erklärt sich aber aus der Tatsache, daß die Zufallsvariablen e_i später die Rolle von Fehlervariablen in einem AR(1)-Prozeß einnehmen werden. Um die Konsistenz der Bezeichnungen in der ganzen Arbeit zu gewährleisten, werden sie deshalb bereits hier mit e_i bezeichnet.

2.1.1 Der empirische Prozeß mit geschätztem Parameter

Eine statistische Fragestellung in diesem Kontext ist, ob F – falls unbekannt – gleich einer fest vorgegebenen Verteilungsfunktion F_0 ist oder nicht. Dies entspricht einem Test der Hypothese

$$\mathrm{H}_0 : \mathrm{F} = \mathrm{F}_0$$

gegen die Alternative

$$\mathrm{H}_1 : \mathrm{F} \neq \mathrm{F}_0.$$

Ein berühmter Test für dieses Problem ist der Kolmogorov-Smirnov-Test. Dabei geht man wie folgt vor: Man schätzt für jedes $x \in \mathbb{R}$ die wahre unbekannte Verteilungsfunktion F an der Stelle x durch

$$\mathrm{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{e_i \le x\}},$$

die *empirische Verteilungsfunktion* auf Basis von $e_1, \ldots, e_n$. Die so definierte Funktion ist nach dem Satz von Glivenko-Cantelli ein gleichmäßig stark konsistenter Schätzer für das wahre F, das heißt es gilt

$$\sup_{x \in \mathbb{R}} |\mathrm{F}_n(x) - \mathrm{F}(x)| \xrightarrow{fs} 0 \qquad \text{für } n \to \infty.$$

Darum ist es von Interesse, das Verteilungsverhalten des empirischen Prozesses $(n^{1/2}(\mathrm{F}_n - \mathrm{F}_0))_{n\in\mathbb{N}}$ zu studieren. Neben Untersuchungen für festes n wurde insbesondere das asymptotische Verteilungsverhalten dieses Prozesses analysiert. Das Hauptresultat in diesem Zusammenhang ist der Satz von Donsker, welcher besagt, daß unter der Hypothese $\mathrm{H}_0 : \mathrm{F} = \mathrm{F}_0$ die Aussage

$$\sqrt{n}\,(\mathrm{F}_n - \mathrm{F}_0) \xrightarrow{\mathcal{L}} \mathrm{B}^0\,(\mathrm{F}_0) \qquad \text{in } \mathrm{D}[-\infty,\infty] \qquad \text{für } n \to \infty \tag{2.1}$$

gilt. Dabei bezeichnet B^0 die Brown'sche Brücke auf $[0,1]$, $\mathrm{D}[-\infty,\infty]$ den Skorohodraum zum Zeitintervall $[-\infty,\infty]$ und $\xrightarrow{\mathcal{L}}$ die Verteilungskonvergenz bezüglich der Skorohodmetrik auf $[-\infty,\infty]$. Diese Verteilungskonvergenzaussage für den empirischen Prozeß kann man mit Hilfe des Stetigkeitssatzes auf die Kolmogorov-Smirnov-Statistik $\sup_{x\in\mathbb{R}} |n^{1/2}(\mathrm{F}_n(x) - \mathrm{F}_0(x))|$ übertragen und erhält so asymptotische kritische Werte des Kolmogorov-Smirnov-Tests für das obige Testproblem.

Diese Fragestellung ist jedoch durch die Festlegung auf eine einzige Verteilungsfunktion F_0 sehr einschränkend und deshalb für die Praxis wenig relevant. Viel stärker ist man an der Frage interessiert, ob F zu einer bestimmten Klasse $\mathcal{F}$ von Verteilungen gehört oder nicht. In diesem Fall möchte man also

$$\mathrm{H}_0 : \mathrm{F} \in \mathcal{F} \qquad \text{gegen} \qquad \mathrm{H}_1 : \mathrm{F} \notin \mathcal{F} \tag{2.2}$$

testen.

Interessierende Klassen $\mathcal{F}$ sind zum Beispiel die Klasse aller Normalverteilungen, also

$$\mathcal{F} = \left\{\mathcal{N}(\mu,\sigma^2) | (\mu,\sigma^2) \in \mathbb{R} \times (0,\infty)\right\}$$

oder allgemeiner endlichdimensional parametrisierte Klassen der Form

$$\mathcal{F} = \left\{\mathrm{F}(\cdot,\vartheta) | \vartheta \in \Theta \subset \mathbb{R}^d\right\}. \tag{2.3}$$

In (2.3) und im Folgenden sind Vektoren in $\mathbb{R}^d$ stets als Spaltenvektoren zu lesen, und alle folgenden Betrachtungen werden unter der Hypothese des Tests (2.2) durchgeführt, das heißt, daß die zugrunde liegende Verteilung in $\mathcal{F}$ liegt. Jedoch ist sie in diesem Fall auch unter der Hypothese nicht vollständig bekannt, da sie von dem (unbekannten) Parameter ϑ abhängt. Wir bezeichnen den wahren Parameter mit ϑ_0 und schreiben von nun an $\mathrm{F}(\cdot, \vartheta_0)$ anstatt F. Der unbekannte Parameter ϑ_0 muß durch einen geeigneten Schätzer $\hat{\vartheta}_n$ geschätzt werden. Diese Tatsache motiviert die Analyse des Prozesses

$$\sqrt{n}\left(\mathrm{F}_n - \mathrm{F}(\cdot, \hat{\vartheta}_n)\right)_{n\in\mathbb{N}},$$

im Folgenden der *empirische Prozeß mit geschätztem Parameter* genannt.

Der empirische Prozeß mit geschätztem Parameter wurde in der statistischen Literatur intensiv untersucht. Einen dem Satz (2.1) von Donsker entsprechenden funktionalen Grenzwertsatz für diesen Prozeß findet man bei Durbin (1973). Um diesen funktionalen Grenzwertsatz zu erhalten, sind allerdings anders als im Kontext des gewöhnlichen empirischen Prozesses einige Regularitätsbedingungen notwendig. Eine Kollektion hinreichender Bedingungen wird im Folgenden aufgelistet.

Es ist $(e_i)_{i\in\mathbb{N}}$ eine Folge unabhängig und identisch nach $\mathrm{F}(\cdot, \vartheta_0)$ verteilter Zufallsvariablen mit $\vartheta_0 \in \Theta$, und die Verteilungsfunktion $\mathrm{F}(\cdot, \vartheta_0)$ ist stetig. (2.4)

Es ist ϑ_0 ein innerer Punkt des Parameterraums Θ, und es existiert eine stetige Funktion $\Delta(\cdot, \vartheta_0) : \mathbb{R} \longrightarrow \mathbb{R}^d$, so daß

$$\sup_{x\in\mathbb{R}} |\mathrm{F}(x, \vartheta_0 + h) - \mathrm{F}(x, \vartheta_0) - \Delta(x, \vartheta_0)^t h| = o(\|h\|) \qquad \text{für } \|h\| \to 0 \tag{2.5}$$

gilt. Hierbei bezeichnet y^t das Transponieren des Vektors $y \in \mathbb{R}^d$ und $\|\cdot\|$ die euklidische Norm auf $\mathbb{R}^d$. Natürlich ist an dieser Stelle auch jede andere Norm auf $\mathbb{R}^d$ geeignet.

Es existiert eine meßbare Funktion $\mathrm{L}(\cdot, \vartheta_0) : \mathbb{R} \longrightarrow \mathbb{R}^d$ mit Erwartungswertvektor $\mathrm{E}(\mathrm{L}(e_1, \vartheta_0)) = 0$ und existierender Kovarianzmatrix $\Lambda(\vartheta_0)$, also $\Lambda(\vartheta_0)_{kl} = \mathrm{E}(\mathrm{L}(e_1, \vartheta_0)\mathrm{L}(e_1, \vartheta_0)^t)_{kl} < \infty$ für alle $1 \le k, l \le d$ und

$$\hat{\vartheta}_n - \vartheta_0 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{L}(e_i, \vartheta_0) + R_n. \tag{2.6}$$

Dabei gilt $\|R_n\| = o_{\mathrm{P}}(n^{-1/2})$.

Die Funktion $\mathrm{L}(\cdot, \vartheta_0)$ in (2.6) heißt *Einflußfunktion*. Die stochastische Entwicklung in (2.6) impliziert über den multivariaten zentralen Grenzwertsatz die Konvergenz von $(n^{1/2}(\hat{\vartheta}_n - \vartheta_0))_{n\in\mathbb{N}}$ gegen eine zentrierte d-dimensionale Normalverteilung mit

Kovarianzmatrix $\Lambda(\vartheta_0)$. In vielen Modellen $\mathcal{F}$ ist diese Entwicklung für zum Beispiel den Maximum-Likelihood-Schätzer für ϑ_0 möglich. Insofern ist (2.6) eine natürliche Bedingung.

Die Aussage (2.5) ist gerade die mehrdimensionale Differenzierbarkeit der Abbildung $\vartheta \longmapsto \mathrm{F}(\cdot, \vartheta)$, und für alle $x \in \mathbb{R}$ ist

$$\Delta(x, \vartheta_0) = \text{grad } \mathrm{F}(x, \vartheta_0) = \left(\frac{\partial}{\partial \vartheta_1} \mathrm{F}(x, \vartheta_0), \ldots, \frac{\partial}{\partial \vartheta_d} \mathrm{F}(x, \vartheta_0) \right)^t,$$

wobei $\frac{\partial}{\partial \vartheta_k} \mathrm{F}(x, \vartheta_0)$ für $k \in \{1, \ldots, d\}$ die partielle Ableitung von $\mathrm{F}(x, \vartheta)$ nach ϑ_k, der k-ten Komponente von ϑ, an der Stelle ϑ_0 bezeichnet. Weiter erhalten wir für alle $k \in \{1, \ldots, d\}$, daß

$$\lim_{x \to \pm\infty} \Delta(x, \vartheta_0)_k = \lim_{x \to \pm\infty} \frac{\partial}{\partial \vartheta_k} \mathrm{F}(x, \vartheta_0) = 0$$

gilt, was insbesondere

$$\lim_{x \to \pm\infty} \|\Delta(x, \vartheta_0)\| = 0 \tag{2.7}$$

impliziert. Dies folgt aus (2.5), wenn wir beachten, daß für alle $\vartheta \in \Theta$ die Funktion $\mathrm{F}(\cdot, \vartheta)$ eine Verteilungsfunktion ist und somit für alle $\vartheta, \vartheta' \in \Theta$ die Aussage

$$\lim_{x \to \pm\infty} \mathrm{F}(x, \vartheta) - \mathrm{F}(x, \vartheta') = 0$$

gilt.

Wir wenden uns nun dem funktionalen Grenzwertsatz für den empirischen Prozeß mit geschätztem Parameter von Durbin zu.

Satz 2.1 *(vgl. Satz 1 aus Durbin (1973))*

Unter den Voraussetzungen (2.4)-(2.6) gilt:

$$\sqrt{n}(\mathrm{F}_n - \mathrm{F}(\cdot, \hat{\vartheta}_n)) \xrightarrow{\mathcal{L}} \mathrm{Z} \qquad \text{in } \mathrm{D}[-\infty, \infty] \qquad \text{für } n \to \infty. \tag{2.8}$$

Dabei ist der Prozeß Z ein zentrierter Gaußprozeß mit stetigen Pfaden. Dieser hat die Kovarianzfunktion

$$\begin{aligned} \mathrm{Cov}\,(\mathrm{Z}(x), \mathrm{Z}(y)) = {} & \mathrm{F}\,(\min\{x, y\}, \vartheta_0) - \mathrm{F}(x, \vartheta_0)\mathrm{F}(y, \vartheta_0) \\ & -\Delta(x, \vartheta_0)^t \mathrm{E}\left(\mathrm{L}(e_1, \vartheta_0) 1_{\{e_1 \le y\}}\right) - \Delta(y, \vartheta_0)^t \mathrm{E}\left(\mathrm{L}(e_1, \vartheta_0) 1_{\{e_1 \le x\}}\right) \\ & +\Delta(x, \vartheta_0)^t \Lambda(\vartheta_0) \Delta(y, \vartheta_0) \qquad \text{für } x, y \in \mathbb{R}. \end{aligned} \tag{2.9}$$

Die von Durbin bewiesene Version ist eine Konvergenzaussage in $\mathrm{D}[0, 1]$, dem Skorohodraum zum Zeitintervall auf $[0, 1]$, diese kann aber mittels einer Zeittransformation von $[0, 1]$ auf $[-\infty, \infty]$ in obige Form überführt werden.

In Abschnitt 2.2 werden wir sehen, wie Durbins funktionaler Grenzwertsatz modifiziert werden kann, wenn man zusätzlich die Zentriertheit von $\mathrm{F}(\cdot, \vartheta_0)$ voraussetzt. In diesem Fall sollte man die Zusatzinformation $\mathrm{E}(e_1) = 0$ natürlich in die Schätzung von $\mathrm{F}(\cdot, \vartheta_0)$ einbeziehen.

2.1.2 Schätzung unter Ausnutzung der Zentriertheit

Während im letzten Abschnitt noch keinerlei Integrierbarkeitsvoraussetzungen an $\mathrm{F}(\cdot, \vartheta_0)$ gestellt werden mußten, wollen wir im Folgenden annehmen, daß die e_i zentriert sind. Auch die Existenz zweiter Momente soll angenommen werden, so daß wir insgesamt also neben den Voraussetzungen aus Abschnitt 2.1.1 auch

$$\mathrm{E}\,(e_1) = 0 \qquad \text{und} \qquad 0 < \sigma^2 := \mathrm{Var}\,(e_1) = \mathrm{E}\left(e_1^2\right) < \infty \tag{2.10}$$

voraussetzen. Die zweite Momentenbedingung muß allerdings nicht unbedingt eine weitere Einschränkung des Modells bedeuten, da sie bereits implizit in den Bedingungen (2.6) an die Einflußfunktion enthalten sein kann.

Ist zum Beispiel $\mathcal{F} = \{\mathcal{N}\,(0, \sigma^2) : \sigma^2 \in (0, \infty)\}$ die Klasse aller zentrierten Normalverteilungen, so ergibt sich als Maximum-Likelihood-Schätzer für die wahre Varianz σ_0^2 gerade

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2.$$

Die zugehörige Einflußfunktion ist dann $\mathrm{L}(x, \vartheta_0) = x^2 - \sigma_0^2$, und $\mathrm{E}(\mathrm{L}^2(e_1, \sigma_0^2)) < \infty$ erfordert in diesem Modell sogar die Existenz vierter Momente der e_i.

Die empirische Verteilungsfunktion F_n nutzt die in der Voraussetzung $\mathrm{E}(e_1) = 0$ liegende Information nicht aus, da sie im allgemeinen nicht zentriert ist. Eine Möglichkeit, diese Zusatzinformation in die Konstruktion eines Schätzers $\mathrm{F}_{n,z}$ für $\mathrm{F}(\cdot, \vartheta_0)$ einzubeziehen, bietet das *empirische Likelihood-Prinzip*, im Folgenden kurz EL-Prinzip genannt. Dieses Verfahren zur nichtparametrischen Verteilungsschätzung unter Zusatzinformationen geht auf Owen (1990) zurück und wurde unter anderem von Qin und Lawless (1994) sowie Zhang (1997) weiterentwickelt.

In unserer Situation der Zusatzinformation $\mathrm{E}(e_1) = 0$ betrachtet man dabei alle Schätzer für $\mathrm{F}(\cdot, \vartheta_0)$ von der Form

$$\hat{\mathrm{F}}_n(x) = \sum_{i=1}^{n} p_i 1_{\{e_i \leq x\}} \qquad \text{für } x \in \mathbb{R}$$

mit

$$0 \leq p_1, \ldots, p_n \leq 1 \qquad \text{und} \qquad \sum_{i=1}^{n} p_i = 1$$

und bestimmt die $\mathrm{F}_{n,z}$ definierenden p_i als Maximalstelle der *empirischen Likelihood-Funktion*

$$\mathrm{ELF}(p_1 \ldots, p_n) = \prod_{i=1}^{n} p_i$$

unter der Nebenbedingung

$$0 = \int_{\mathrm{I\!R}} x \hat{\mathrm{F}}_n(dx) = \sum_{i=1}^{n} p_i e_i. \tag{2.11}$$

Durch die Nebenbedingung wird die Zentriertheit aller $\hat{\mathrm{F}}_n$ und damit von $\mathrm{F}_{n,z}$ erzwungen. Das Maximierungsproblem läßt sich mit Hilfe der Methode der Lagrange-Multiplikatoren leicht behandeln und hat die folgende Lösung: Unter der Voraussetzung

$$\min_{1 \le i \le n} e_i < 0 < \max_{1 \le i \le n} e_i \tag{2.12}$$

hat die Gleichung

$$\sum_{i=1}^{n} \frac{e_i}{1 + t_n e_i} = 0$$

im Intervall

$$\left(\left(\frac{1}{n} - 1 \right) \frac{1}{\max_{1 \le i \le n} e_i}, \left(\frac{1}{n} - 1 \right) \frac{1}{\min_{1 \le i \le n} e_i} \right)$$

genau eine Lösung t_n und

$$\hat{p}_{ni} = \frac{1}{n} \frac{1}{1 + t_n e_i} \qquad \text{für } 1 \le i \le n$$

ist die eindeutig bestimmte Maximalstelle von ELF unter der Nebenbedingung (2.11). Damit ist

$$\mathrm{F}_{n,z}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + t_n e_i} 1_{\{e_i \le x\}} \qquad \text{für } x \in \mathrm{I\!R}$$

der EL-Schätzer für $\mathrm{F}(\cdot, \vartheta_0)$ unter der Zusatzinformation $\mathrm{E}(e_1) = 0$. Wie oben bereits erwähnt ist $\mathrm{F}_{n,z}$ gemäß Konstruktion zentriert.

Der Schätzer $\mathrm{F}_{n,z}$ ist eine Spezialisierung des von Qin und Lawless (1994) und Zhang (1997) betrachteten Schätzers (dort wurden auch andere Zusatzinformationen als $\mathrm{E}(e_1) = 0$ zugelassen). Einige wichtige Ergebnisse seien im Folgenden genannt.

Lemma 2.2 *(vgl. Lemma 2 aus Owen (1990))*

Unter der Voraussetzung (2.10) gilt

$$\mathrm{P}\left(0 \notin (\min_{1 \le i \le n} e_i, \max_{1 \le i \le n} e_i) \right) \longrightarrow 0 \qquad \text{für } n \to \infty.$$

Für asymptotische Ergebnisse spielt also dieses Ereignis keine Rolle. Wir werden deshalb in Zukunft immer annehmen, daß (2.12) erfüllt und $\mathrm{F}_{n,z}$ damit wohldefiniert ist.

Lemma 2.3 *(vgl. Beweis zu Satz 1 aus Owen (1990))*

Gilt (2.10), so gilt auch

$$t_n = \frac{1}{\sigma^2} \cdot \frac{1}{n} \sum_{i=1}^{n} e_i + o_{\mathrm{P}}\left(n^{-1/2}\right). \tag{2.13}$$

Diese stochastische Entwicklung erlaubt also die Darstellung von t_n als normierte Summe unabhängig und identisch verteilter Zufallsvariablen plus einem für Verteilungskonvergenzaussagen vernachlässigbaren Fehlerterm.

Lemma 2.4 *(vgl. Satz 3.2 aus Zhang (1997))*

Unter der Bedingung (2.10) gilt für alle $x \in \mathbb{R}$ die Entwicklung

$$\mathrm{F}_{n,z}(x) - \mathrm{F}(x, \vartheta_0) = \frac{1}{n} \sum_{i=1}^{n} \left(1_{\{e_i \leq x\}} - \mathrm{F}(x, \vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2} e_i \right) + R_n(x). \tag{2.14}$$

Hierbei ist

$$\mathrm{U}(x) = \mathrm{E}\left(e_1 1_{\{e_1 \leq x\}}\right) \qquad \text{für } x \in \mathbb{R} \tag{2.15}$$

gesetzt, und es gilt

$$\sup_{x \in \mathbb{R}} |R_n(x)| = o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Diese stochastische Entwicklung von $\mathrm{F}_{n,z}$ erlaubt die Herleitung eines funktionalen Grenzwertsatzes für den Prozeß $(n^{1/2}(\mathrm{F}_{n,z} - \mathrm{F}(\cdot, \vartheta_0)))_{n \in \mathbb{N}}$; vgl. Satz 3.3 in Zhang (1997).

Einige der oben genannten Aussagen werden von den Autoren unter stärkeren Momentenvoraussetzungen als $\mathrm{Var}(e_1) < \infty$ bewiesen, da sie allgemeinere Fragestellungen untersuchen. Wegen der Spezialisierung auf die eine Zusatzbedingung $\mathrm{E}(e_1) = 0$ können diese Momentenbedingungen aber zu (2.10) abgeschwächt werden; vgl. Genz (2001).

2.2 Der modifizierte empirische Prozeß mit geschätztem Parameter

In diesem Abschnitt diskutieren wir das asymptotische Verhalten des *modifizierten empirischen Prozesses mit geschätztem Parameter* $(n^{1/2}(\mathrm{F}_{n,z} - \mathrm{F}(\cdot, \hat{\vartheta}_n)))_{n \in \mathbb{N}}$. Dabei erhalten wir zunächst das folgende zu Lemma 2.4 analoge Lemma:

Lemma 2.5

Unter den Bedingungen (2.4)-(2.6) und (2.10) gilt für alle $x \in \mathbb{R}$ die Aussage

$$\mathrm{F}_{n,z}(x) - \mathrm{F}(x, \hat{\vartheta}_n) = \frac{1}{n} \sum_{i=1}^{n} Y_i(x) + R_n(x) \tag{2.16}$$

mit

$$Y_i(x) = 1_{\{e_i \leq x\}} - \mathrm{F}(x, \vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2} e_i - \Delta(x, \vartheta_0)^t \mathrm{L}\left(e_i, \vartheta_0\right). \tag{2.17}$$

Dabei ist U wieder gemäß (2.15) definiert, und für den Restterm gilt

$$\sup_{x \in \mathbb{R}} |R_n(x)| = o_{\mathrm{P}}\left(n^{-1/2}\right). \tag{2.18}$$

Beweis Zunächst gilt für alle $x \in \mathbb{R}$ die Gleichung

$$\begin{aligned} \mathrm{F}_{n,z}(x) - \mathrm{F}(x, \hat{\vartheta}_n) &= (\mathrm{F}_{n,z}(x) - \mathrm{F}(x, \vartheta_0)) - (\mathrm{F}(x, \hat{\vartheta}_n) - \mathrm{F}(x, \vartheta_0)) \\ &= A_{1n}(x) - A_{2n}(x). \end{aligned}$$

Gemäß (2.4) und (2.10) können wir die Darstellung (2.14) für $A_{1n}(x)$ verwenden, das heißt wir können $A_{1n}(x)$ wie folgt umformen:

$$A_{1n}(x) = \frac{1}{n} \sum_{i=1}^{n} \left(1_{\{e_i \leq x\}} - \mathrm{F}(x, \vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2} e_i \right) + R_{1n}(x).$$

Dabei hat der Restterm R_{1n} bereits die gewünschte Ordnung.

Es bleibt A_{2n} zu untersuchen. Die Bedingung (2.6) impliziert insbesondere die Konsistenz von $\hat{\vartheta}_n$. Damit liegt der Schätzer $\hat{\vartheta}_n$ mit gegen Eins konvergierender Wahrscheinlichkeit in einer Umgebung von ϑ_0. Nun gilt aber

$$\begin{aligned} A_{2n}(x) &= \Delta(x, \vartheta_0)^t \left(\hat{\vartheta}_n - \vartheta_0 \right) \\ &\quad + \left(\mathrm{F}(x, \hat{\vartheta}_n) - \mathrm{F}(\cdot, \vartheta_0) - \Delta(x, \vartheta_0)^t \left(\hat{\vartheta}_n - \vartheta_0 \right) \right) \\ &= \Delta(x, \vartheta_0)^t \left(\hat{\vartheta}_n - \vartheta_0 \right) + R_{2n}(x), \end{aligned}$$

und (2.5) und (2.6) implizieren

$$\sup_{x \in \mathbb{R}} |R_{2n}(x)| = o(\|\hat{\vartheta}_n - \vartheta_0\|) = o_{\mathrm{P}}(n^{-1/2}).$$

Andererseits erhalten wir mittels der stochastischen Entwicklung (2.6) für alle $x \in \mathbb{R}$ die Gleichung

$$\begin{aligned} \Delta(x, \vartheta_0)^t \left(\hat{\vartheta}_n - \vartheta_0 \right) &= \Delta(x, \vartheta_0)^t \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}\left(e_i, \vartheta_0\right) + \Delta(x, \vartheta_0)^t \cdot R_n \\ &= \Delta(x, \vartheta_0)^t \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}\left(e_i, \vartheta_0\right) + R_{3n}(x) \end{aligned}$$

Dabei ist $\|R_n\| = o_{\mathrm{P}}(n^{-1/2})$, und mit Hilfe der Cauchy-Schwarz-Ungleichung erhalten wir

$$\sup_{x\in\mathbb{R}} |R_{3n}(x)| \leq \sup_{x\in\mathbb{R}} \|\Delta(x,\vartheta_0)\| \cdot \|R_n\| = o_{\mathrm{P}}(n^{-1/2}),$$

denn wegen (2.5) ist $\Delta(\cdot,\vartheta_0)$ stetig und mit (2.7) ist somit $\|\Delta(\cdot,\vartheta_0)\|$ beschränkt.

Schließlich folgern wir die Behauptung, indem wir $R_n(x) = R_{1n}(x) - R_{2n}(x) - R_{3n}(x)$ setzen. □

Die stochastische Entwicklung (2.14) erlaubt uns nun, einen zu (2.8) analogen funktionalen Grenzwertsatz für den stochastischen Prozeß $(n^{1/2}(\mathrm{F}_{n,z} - \mathrm{F}(\cdot,\hat{\vartheta}_n)))_{n\in\mathbb{N}}$ zu beweisen. Es gilt:

Satz 2.6

Unter den Voraussetzungen (2.4)-(2.6) und (2.10) gilt

$$\sqrt{n}\left(\mathrm{F}_{n,z} - \mathrm{F}(\cdot,\hat{\vartheta}_n)\right) \xrightarrow{\mathcal{L}} \mathrm{W} \qquad \text{in } \mathrm{D}[-\infty,\infty] \qquad \text{für } n\to\infty. \tag{2.19}$$

Dabei ist W ein zentrierter Gaußprozeß mit stetigen Pfaden und der Kovarianzfunktion

$$\begin{aligned} \mathrm{Cov}\left(\mathrm{W}(x),\mathrm{W}(y)\right) &= \mathrm{F}\left(\min\{x,y\},\vartheta_0\right) - \mathrm{F}(x,\vartheta_0)\mathrm{F}(y,\vartheta_0) - \frac{\mathrm{U}(x)\mathrm{U}(y)}{\sigma^2} \\ &-\Delta(x,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1,\vartheta_0)1_{\{e_1\leq y\}}\right) - \Delta(y,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1,\vartheta_0)1_{\{e_1\leq x\}}\right) \\ &+\frac{\mathrm{U}(x)}{\sigma^2}\Delta(y,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1,\vartheta_0)e_1\right) + \frac{\mathrm{U}(y)}{\sigma^2}\Delta(x,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1,\vartheta_0)e_1\right) \\ &+\Delta(x,\vartheta_0)^t\Lambda(\vartheta_0)\Delta(y,\vartheta_0) \qquad \text{für } x,y\in\mathbb{R}. \end{aligned} \tag{2.20}$$

Beweis Gemäß (2.16) genügt es, die Verteilungskonvergenzaussage für den Prozeß $(n^{-1/2}\sum_{i=1}^n Y_i)_{n\in\mathbb{N}}$ zu zeigen. Dazu wiederum reicht es aus, die Konvergenz der endlichdimensionalen Randverteilungen und die C-Straffheit des Prozesses zu beweisen (vgl. Billingsley (1968) Theorem 15.1 und 15.5, siehe hierzu auch Satz 2.4 aus Genz (2001)).

Zunächst wenden wir uns den endlichdimensionalen Randverteilungen des Prozesses $\left(n^{-1/2}\sum_{i=1}^n Y_i\right)_{n\in\mathbb{N}}$ zu. Seien hierzu $\{t_1,\ldots,t_m\}$ Punkte in $[-\infty,\infty]$ und

$$\underline{Y}_i(\underline{t}_m) = (Y_i(t_1),\ldots,Y_i(t_m))^t.$$

Berechnen wir für $k,l\in\{1,\ldots,m\}$ die Einträge Γ_{kl} der Kovarianzmatrix Γ des Vektors $\underline{Y}_1(\underline{t}_m)$, so ergibt sich durch elmentare Rechnungen, daß

$$\Gamma_{kl} = \mathrm{Cov}(Y_1(t_k),Y_1(t_l)) = \mathrm{Cov}(\mathrm{W}(t_k),\mathrm{W}(t_l))$$

gilt, und da $n^{-1/2}\sum_{i=1}^n \underline{Y}_i(\underline{t}_m)$ eine Summe unabhängig und identisch verteilter zentrierter Zufallsvektoren ist, liefert uns der mehrdimensionale zentrale Grenzwertsatz, daß

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \underline{Y}_i(\underline{t}_m) \xrightarrow{\mathcal{L}} \mathcal{N}_m\left(0,\Gamma\right) \qquad \text{für } n\to\infty \tag{2.21}$$

gilt, womit die Verteilungskonvergenz der endlichdimensionalen Randverteilungen bewiesen ist.

Es bleibt die C-Straffheit des Prozesses $(n^{-1/2}\sum_{i=1}^n Y_i)_{n\in\mathbb{N}}$ zu zeigen. Dazu stellen wir zunächst fest, daß für alle $\mathrm{f},\mathrm{g}\in \mathrm{D}[-\infty,\infty]$ und $\delta>0$ die Abschätzung

$$\mathrm{w}_\infty\left(f+g,\delta\right)\le \mathrm{w}_\infty\left(f,\delta\right)+\mathrm{w}_\infty\left(g,\delta\right)$$

richtig ist. Dabei bezeichnet w_∞ den Stetigkeitsmodul auf $\mathrm{D}[-\infty,\infty]$, also

$$\mathrm{w}_\infty(\mathrm{f},\delta)=\sup_{\rho(s,t)\le\delta}|\mathrm{f}(s)-\mathrm{f}(t)|$$

mit einer Metrik ρ zur euklidischen Topologie auf $[-\infty,\infty]$. Es ist nun zu zeigen, daß für alle $\varepsilon>0$ die Aussage

$$\limsup_{n\to\infty}\mathrm{P}\left(\mathrm{w}_\infty\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i,\delta\right)\ge\varepsilon\right)\longrightarrow 0 \qquad \text{für } \delta\to 0 \tag{2.22}$$

gilt. Sei also $\varepsilon>0$. Dann gilt

$$\begin{aligned}
&\limsup_{n\to\infty}\mathrm{P}\left(\mathrm{w}_\infty\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i,\delta\right)\ge\varepsilon\right)\\
\le\ &\limsup_{n\to\infty}\mathrm{P}\left(\mathrm{w}_\infty\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(1_{\{e_i\le\cdot\}}-\mathrm{F}(\cdot,\vartheta_0)-\frac{\mathrm{U}}{\sigma^2}e_i\right),\delta\right)\ge\frac{\varepsilon}{2}\right)\\
&+\limsup_{n\to\infty}\mathrm{P}\left(\mathrm{w}_\infty\left(\Delta(\cdot,\vartheta_0)^t\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathrm{L}(e_i,\vartheta_0),\delta\right)\ge\frac{\varepsilon}{2}\right).
\end{aligned}$$

Der erste Summand auf der rechten Seite der Ungleichung wurde bereits in Genz (2001) untersucht. Dort konnte gezeigt werden, daß er für $\delta\to 0$ gegen Null konvergiert. Es existiert also ein $\delta_1>0$, so daß für alle $0<\delta<\delta_1$ die Ungleichung

$$\limsup_{n\to\infty}\mathrm{P}\left(\mathrm{w}_\infty\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(1_{\{e_i\le\cdot\}}-\mathrm{F}(\cdot,\vartheta_0)-\frac{\mathrm{U}}{\sigma^2}e_i\right),\delta\right)\ge\frac{\varepsilon}{2}\right)\le\frac{\varepsilon}{2} \tag{2.23}$$

gilt. Es bleibt folglich lediglich der zweite Summand zu untersuchen. Da (2.6) insbesondere

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathrm{L}\left(e_i,\vartheta_0\right)\right\|=O_{\mathrm{P}}(1)$$

impliziert, existiert ein $C > 0$, so daß

$$\limsup_{n\to\infty} \mathrm{P}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{L}\left(e_i,\vartheta_0\right)\right)\right\| \geq C\right) \leq \frac{\varepsilon}{2}$$

gilt. Andererseits ist $\Delta(\cdot,\vartheta_0)$ wegen (2.5) und (2.7) gleichmäßig stetig, weshalb wir $0 < \delta_2 < \delta_1$ so wählen können, daß

$$\|\Delta(x,\vartheta_0) - \Delta(y,\vartheta_0)\| \leq \frac{\varepsilon}{4C} \qquad \text{für } |x-y| \leq \delta_2$$

gilt. Insgesamt ergibt sich also für alle $0 < \delta < \delta_2$ die Ungleichung

$$\begin{aligned}
&\limsup_{n\to\infty} \mathrm{P}\left(\mathrm{w}_\infty\left(\Delta(\cdot,\vartheta_0)^t\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{L}\left(e_i,\vartheta_0\right),\delta\right) \geq \frac{\varepsilon}{2}\right)\\
\leq\ &\limsup_{n\to\infty} \mathrm{P}\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{L}\left(e_i,\vartheta_0\right)\right\| \geq C\right)\\
\leq\ &\frac{\varepsilon}{2},
\end{aligned}$$

und zusammen mit (2.23) erhalten wir (2.22), das ist die C-Straffheit des Prozesses $(n^{-1/2}\sum_{i=1}^{n} Y_i)_{n\in\mathbb{N}}$. □

Zum Abschluß dieses Kapitels soll noch herausgestellt werden, inwiefern der gerade bewiesene funktionale Grenzwertsatz eine Verbesserung der Aussage (2.8) bedeutet. Hierzu vergleichen wir für jedes $x \in \mathbb{R}$ die Varianzen der Grenzprozesse Z und W. Dabei stellt sich heraus, daß

$$\mathrm{Var}(\mathrm{Z}(x)) - \mathrm{Var}(\mathrm{W}(x)) = \frac{\mathrm{U}(x)^2}{\sigma^2} - 2\frac{U(x)}{\sigma^2}\Delta(x,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1,\vartheta_0)e_1\right)$$

gilt. Im allgemeinen kann über diese Differenz keine quantitative Aussage gemacht werden. Sie ist aber positiv, falls

$$\mathrm{E}(e_1\mathrm{L}(e_1,\vartheta_0)) = 0 \tag{2.24}$$

gilt. Dies bedeutet, daß der Prozeß W in geeigneten Verteilungsmodellen eine punktweise nichtgrößere Varianz als der Prozeß Z besitzt. So ist im oben angesprochenen Beispiel der Klasse der zentrierten Normalverteilungen

$$\mathrm{E}(\mathrm{L}(e_1,\sigma_0^2)e_1) = \mathrm{E}(e_1^3 - \sigma_0^2 e_1) = 0$$

und die Differenz der Varianzen

$$\mathrm{Var}(\mathrm{Z}(x)) - \mathrm{Var}(\mathrm{W}(x)) = \frac{\mathrm{U}(x)^2}{\sigma_0^2} \geq 0.$$

Dabei gilt nicht für alle $x \in \mathbb{R}$ Gleichheit.

Betrachten wir allgemeiner den Fall, daß ϑ_0 ein beliebiger Verteilungsparameter ist, wobei wir uns zur Notationsvereinfachung zunächst auf den eindimensionalen Fall $\vartheta_0 \in \Theta \subset \mathbb{R}$ beschränken. Dann gilt der folgende allgemeine Satz über Maximum-Likelihood-Schätzer für ϑ_0.

Satz 2.7 *(vgl. z.B. Satz 2.32 bei Witting und Nölle (1970))*

Sei ϑ_0 ein innerer Punkt von $\Theta \subset \mathrm{I\!R}$, für alle $\vartheta \in \Theta$ besitze $\mathrm{F}(\cdot, \vartheta)$ eine $\lambda - f\ddot{u}$ positive Dichte $\mathrm{f}(\cdot, \vartheta)$ und es gelten die Regularitätsbedingungen:

(i) Es existiert eine offene Umgebung V von ϑ_0, so daß f für alle $(x, \vartheta) \in \mathrm{I\!R} \times V$ zweimal stetig partiell differenzierbar nach ϑ ist.

(ii) Es gilt $0 < \mathrm{I}(\vartheta_0) = \mathrm{E}\left(\left(\frac{\partial}{\partial\vartheta} \log \mathrm{f}(e_1, \vartheta_0)\right)^2\right) < \infty$.

(iii) Es ist

$$\mathrm{E}\left(\frac{1}{\mathrm{f}(e_1, \vartheta_0)} \frac{\partial}{\partial\vartheta} \mathrm{f}(e_1, \vartheta_0)\right) = 0 \qquad \text{und} \qquad \mathrm{E}\left(\frac{1}{\mathrm{f}(e_1, \vartheta_0)} \frac{\partial^2}{\partial\vartheta^2} \mathrm{f}(e_1, \vartheta_0)\right) = 0.$$

(iv) Es existiert eine Funktion $\mathrm{M}_{\vartheta_0} : \mathrm{I\!R} \longrightarrow [0, \infty)$ mit

$$\int_{\mathrm{I\!R}} \mathrm{M}_{\vartheta_0}(x) \mathrm{F}(dx, \vartheta_0) < \infty$$

für die für alle ϑ in V und $x \in \mathrm{I\!R}$ gilt:

$$\left| \frac{\partial^2}{\partial\vartheta^2} \log \mathrm{f}(x, \vartheta) \right| \leq \mathrm{M}_{\vartheta_0}(x).$$

Dann hat jeder konsistente Maximum-Likelihood-Schätzer $\hat{\vartheta}_n$ für ϑ_0 eine stochastische Entwicklung gemäß (2.6) mit der Einflußfunktion

$$\mathrm{L}(x, \vartheta_0) = \frac{1}{\mathrm{I}(\vartheta_0)} \frac{\partial}{\partial\vartheta} \log \mathrm{f}(x, \vartheta_0).$$

Eine hinreichende Bedingung für (iii) ist zum Beispiel die Existenz einer Funktion $\mathrm{H}_{\vartheta_0} : \mathrm{I\!R} \longrightarrow [0, \infty)$ mit

$$\int_{-\infty}^{\infty} \mathrm{H}_{\vartheta_0}(x) dx < \infty,$$

die

$$\left| \frac{\partial}{\partial\vartheta} \mathrm{f}(x, \vartheta) \right| \leq \mathrm{H}_{\vartheta_0}(x) \qquad \text{und} \qquad \left| \frac{\partial^2}{\partial\vartheta^2} \mathrm{f}(x, \vartheta) \right| \leq \mathrm{H}_{\vartheta_0}(x) \qquad \text{für } \vartheta \in V, x \in \mathrm{I\!R}$$

erfüllt. Sind nun die $\mathrm{f}(\cdot, \vartheta)$ zentriert und gilt zusätzlich

$$\int_{-\infty}^{\infty} |x| \mathrm{H}_{\vartheta_0}(x) dx < \infty,$$

so folgt

$$\mathrm{E}\left(e_1\mathrm{L}(e_1,\vartheta_0)\right) = \frac{1}{\mathrm{I}(\vartheta_0)}\int_{-\infty}^{\infty} x\frac{\partial}{\partial\vartheta}\mathrm{f}(x,\vartheta_0)dx = \frac{1}{\mathrm{I}(\vartheta_0)}\frac{\partial}{\partial\vartheta}\int_{-\infty}^{\infty} x\mathrm{f}(x,\vartheta)dx\bigg|_{\vartheta=\vartheta_0} = 0,$$

wobei das letzte Gleichheitszeichen wegen der Zentriertheit der $\mathrm{f}(\cdot,\vartheta)$ gilt. Die Vertauschbarkeit von Integral und partieller Ableitung nach ϑ liefert der Integralkonvergenzsatz von Lebesgue, da der Integrand lokal gleichmäßig um ϑ_0 durch die integrierbare Funktion $\mathbb{R} \ni x \longmapsto |x|\mathrm{H}_{\vartheta_0}(x)$ beschränkt ist.

Insgesamt sehen wir damit, daß wir unter natürlichen Regularitätsbedingungen an die Verteilungsklasse $\mathcal{F}$ für den Maximum-Likelihood-Schätzer für ϑ_0 gerade $\mathrm{E}(e_1\mathrm{L}(e_1,\vartheta_0)) = 0$ und damit

$$\mathrm{Var}(\mathrm{Z}(x)) - \mathrm{Var}(\mathrm{W}(x)) = \frac{\mathrm{U}(x)^2}{\sigma^2} \geq 0 \tag{2.25}$$

erhalten. Für $t_1 \ldots, t_m \in \mathbb{R}$ gilt sogar allgemeiner

$$\mathrm{Cov}\left((\mathrm{Z}(t_1),\ldots,\mathrm{Z}(t_m))^t\right) - \mathrm{Cov}\left((\mathrm{W}(t_1),\ldots,\mathrm{W}(t_m))^t\right) = \frac{1}{\sigma^2}\cdot \underline{\mathrm{U}}(\underline{t}_m)\underline{\mathrm{U}}(\underline{t}_m)^t,$$

wobei

$$\underline{\mathrm{U}}(\underline{t}_m) = (\mathrm{U}(t_1)\ldots,\mathrm{U}(t_m))^t$$

gesetzt ist. Betrachtet man also die auf dem Raum der symmetrischen Matrizen definierte Loewner-Halbordnung (vgl. z.B. Pukelsheim (1993), Absatz 1.10)

$$A \leq B \qquad \text{genau dann wenn} \qquad B - A \text{ ist nichtnegativ definit,}$$

so gilt gerade

$$\mathrm{Cov}\left((\mathrm{W}(t_1),\ldots,\mathrm{W}(t_m))^t\right) \leq \mathrm{Cov}\left((\mathrm{Z}(t_1),\ldots,\mathrm{Z}(t_m))^t\right),$$

da $\underline{\mathrm{U}}(\underline{t}_m)\underline{\mathrm{U}}(\underline{t}_m)^t$ trivialerweise nichtnegativ definit ist. Es gilt also auch die mehrdimensionale Version von (2.25) für alle endlichdimensionalen Randverteilungen der Prozesse W und Z.

Die vorangegangenen Überlegungen gelten in analoger Form auch für mehrdimensionale Verteilungsparameter.

Wie sich der Unterschied in der Varianz der Grenzprozesse Z und W beim Testen auf die Verteilungsklasse $\mathcal{F} = \left\{\mathrm{F}(\cdot,\vartheta)|\vartheta \in \Theta \subset \mathbb{R}^d\right\}$ auswirkt, werden wir in einem späteren Kapitel noch empirisch im Rahmen einer Simulationsstudie untersuchen.

Kapitel 3

Der empirische Prozeß der Residuen mit geschätztem Parameter bei autoregressiven Prozessen erster Ordnung

3.1 Modellvoraussetzungen und bekannte Ergebnisse

Während wir im letzten Kapitel die Situation unabhängig und identisch verteilter Zufallsvariablen untersucht haben, wollen wir in diesem Kapitel Zufallsvariablen zugrunde legen, die einem autoregressiven Prozeß erster Ordnung, im Folgenden kurz AR(1)-Prozeß oder AR(1)-Zeitreihe genannt, entspringen. Dazu betrachten wir wie im vorangegangenen Kapitel eine Folge $(e_i)_{i\in\mathbb{N}}$ von Zufallsvariablen, die wieder stets unabhängig und identisch nach einer Verteilungsfunktion $\mathrm{F}(\cdot,\vartheta_0) \in \mathcal{F}$ verteilt sein sollen. Weiter sei X_0 eine von $(e_i)_{i\in\mathbb{N}}$ unabhängige Zufallsvariable und ρ ein reeller Parameter. Mithilfe der Rekursion

$$X_n = \rho X_{n-1} + e_n \qquad \text{für } n \geq 1$$

erhalten wir dann den AR(1)-Prozeß $(X_n)_{n\in\mathbb{N}_0}$. Durch mehrfaches Anwenden dieser Formel können wir die X_n explizit durch

$$X_n = \rho^n X_0 + \sum_{i=1}^{n} \rho^{n-i} e_i \qquad \text{für } n \geq 1 \tag{3.1}$$

darstellen. Die so definierten X_n sind im allgemeinen natürlich nicht mehr unabhängig und identisch verteilt, und ihr asymptotisches Verhalten wird im wesentlichen durch $\sum_{i=1}^{n} \rho^{n-i} e_i$ bestimmt. Wir werden deshalb im Folgenden zur technischen

Vereinfachung

$$X_0 = 0$$

annehmen. Die unter dieser Annahme gewonnenen Ergebnisse lassen sich wegen (3.1) mit Routineargumenten auf den Fall beliebiger von $(e_i)_{i\in\mathbb{N}}$ unabhängiger Startwerte X_0 übertragen.

Wie schon im letzten Kapitel setzen wir auch im AR(1)-Modell (2.10) also $\mathrm{E}(e_1) = 0$ und $0 < \sigma^2 = \mathrm{E}(e_1^2) < \infty$ voraus. Dabei ist die vorher betrachtete Zusatzbedingung $\mathrm{E}(e_1) = 0$ im AR(1)-Modell eine natürliche Regularitätsbedingung, wenn die e_i als weißes Rauschen interpretiert werden.

Bei der Analyse des Verteilungsverhaltens von $(X_n)_{n\in\mathbb{N}}$ stellt man fest, daß dem *Autoregressionsparameter* ρ eine wesentliche Rolle zufällt. Man hat den *stabilen* Fall $|\rho| < 1$, den *semistabilen* Fall $|\rho| = 1$ und den *explosiven* Fall $|\rho| > 1$ zu unterscheiden.

3.1.1 Schätzung des Autoregressionsparameters

In statistischen Anwendungen des AR(1)-Modells wollen wir lediglich die Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ als beobachtbar ansehen, während der Regressionsparameter ρ unbekannt ist. Er muß also durch einen Schätzer $\hat{\rho}_n$ auf Basis von $X_0, \ldots, X_n$ geschätzt werden. Über den Parameterschätzer setzen wir im Folgenden stets

$$\hat{\rho}_n - \rho = \begin{cases} O_{\mathrm{P}}(n^{-1/2}) & \text{für } |\rho| < 1 \\ o_{\mathrm{P}}(n^{-1}) & \text{für } |\rho| = 1 \\ O_{\mathrm{P}}(|\rho|^{-n}) & \text{für } |\rho| > 1 \end{cases} \tag{3.2}$$

voraus.

Diese Bedingungen sind insoweit natürlich, als daß sie im stabilen und im explosiven Fall vom Kleinst-Quadrate-Schätzer

$$\tilde{\rho}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}$$

erfüllt werden. Lediglich für den Fall $|\rho| = 1$ ist eine Modifikation des Kleinst-Quadrate-Schätzers nötig, um die gewünschte Konvergenzordnung gegen den wahren Parameter zu erhalten, da der Kleinst-Quadrate-Schätzer hier nur die Ordnung $O_{\mathrm{P}}(n^{-1})$ hat. Eine Möglichkeit dies zu erreichen, ist die Transformation des Kleinst-Quadrate-Schätzers durch eine Folge von Funktionen $\varphi_n : \mathbb{R} \longrightarrow \mathbb{R}$, so daß $\hat{\rho}_n = \varphi_n(\tilde{\rho}_n)$ für alle $\rho \in \mathbb{R}$ die gewünschten Konvergenzordnungen hat. Datta und Sriram (1997) schlagen eine geeignete Funktionenfolge vor. Dabei hat der gewonnene Schätzer nicht nur die gewünschte Konvergenzordnung, auch die φ_n sind auf ganz $\mathbb{R}$ stetig und zufällig, das heißt adaptiv an die Daten angepaßt. Wir wollen hier lediglich die Idee der Modifikation verdeutlichen und geben deshalb eine stark vereinfachte deterministische Version der von Datta und Sriram (1997) verwendeten Funktionen an.

Lemma 3.1 *(vgl. Lemma A.2. aus Datta und Sriram (1997))*

Sei für $n \in \mathbb{N}$ die Funktion $\varphi_n : \mathbb{R} \longrightarrow \mathbb{R}$ definiert durch

$$\varphi_n(x) = \begin{cases} -1 & \text{für } x \in (-1 - n^{-1/2}, -1 + n^{-1/2}) \\ 1 & \text{für } x \in (1 - n^{-1/2}, 1 + n^{-1/2}) \\ x & \text{sonst} \end{cases}$$

und $\hat{\rho}_n = \varphi_n(\tilde{\rho}_n)$. Dann gelten für $\hat{\rho}_n$ die Konvergenzordnungen aus (3.2).

Beweis Sei zunächst $|\rho| \neq 1$ und $\varepsilon = \min\{|\rho - 1|, |\rho + 1|\} > 0$. Dann gilt

$$\begin{aligned} \mathrm{P}\left(\hat{\rho}_n \neq \tilde{\rho}_n\right) &\leq \mathrm{P}\left(|\tilde{\rho}_n + 1| \leq \frac{1}{\sqrt{n}}\right) + \mathrm{P}\left(|\tilde{\rho}_n - 1| \leq \frac{1}{\sqrt{n}}\right) \\ &\leq 2\mathrm{P}\left(|\tilde{\rho}_n - \rho| > \frac{\varepsilon}{2}\right) + \mathrm{P}\left(\left\{|\tilde{\rho}_n + 1| \leq \frac{1}{\sqrt{n}}\right\} \cap \left\{|\tilde{\rho}_n - \rho| \leq \frac{\varepsilon}{2}\right\}\right) \\ &\quad + \mathrm{P}\left(\left\{|\tilde{\rho}_n - 1| \leq \frac{1}{\sqrt{n}}\right\} \cap \left\{|\tilde{\rho}_n - \rho| \leq \frac{\varepsilon}{2}\right\}\right) \\ &= 2\mathrm{P}\left(|\tilde{\rho}_n - \rho| > \frac{\varepsilon}{2}\right) \end{aligned}$$

für alle großen n, und die rechte Seite der Ungleichungskette konvergiert offensichtlich in n gegen Null, da $\tilde{\rho}_n$ ein konsistenter Schätzer für ρ ist. Dies impliziert die gewünschten Konvergenzordnungen für $\hat{\rho}_n$.

Ist andererseits $\rho = 1$, so gilt für alle $\varepsilon > 0$ und alle Folgen $(a_n)_{n\in\mathbb{N}}$ die Ungleichung

$$\mathrm{P}\left(a_n\left|\hat{\rho}_n - 1\right| \geq \varepsilon\right) \leq \mathrm{P}\left(\hat{\rho}_n \neq 1\right) \leq \mathrm{P}\left(n|\tilde{\rho}_n - 1| \geq \sqrt{n}\right),$$

und die letzte Wahrscheinlichkeit konvergiert für $n \to \infty$ gegen Null, was insbesondere die stochastische Konvergenz von $n(\hat{\rho}_n - 1)$ gegen Null beweist. Durch geeignete Wahl der Folge $(a_n)_{n\in\mathbb{N}}$ erhalten wir sogar jede beliebige stochastische Konvergenzordnung. Der Beweis im Fall $\rho = -1$ verläuft analog, womit das Lemma vollständig gezeigt ist. □

Wir haben also die Existenz eines Schätzers $\hat{\rho}_n$ für ρ mit den in (3.2) angegebenen Konvergenzordnungen gezeigt. Wir wenden uns nun der Schätzung von $\mathrm{F}(\cdot, \vartheta_0)$ zu. Dazu stellen wir zunächst fest, daß die Fehler $e_1, \ldots, e_n$ im AR(1)-Modell nicht beobachtbar sind. Um dennoch einen Verteilungsschätzer zu erhalten, verwendet man deshalb zunächst einen beliebigen Schätzer $\hat{\rho}_n$ für ρ zur Definition der *Residuen*

$$\hat{e}_{ni} = X_i - \hat{\rho}_n X_{i-1} \qquad \text{für } 1 \leq i \leq n.$$

Dabei werden wir stets voraussetzen, daß $\hat{\rho}_n$ die Konvergenzordnungen aus (3.2) aufweist.

Obwohl jedes Residuum von $X_0, \ldots, X_n$ und damit von n abhängt, werden wir im Folgenden zur Notationsvereinfachung statt $\hat{e}_{ni}$ nur noch $\hat{e}_i$ schreiben.

3.1.2 Das asymptotische Verhalten der Residuen

Die so konstruierten Residuen $\hat{e}_1, \ldots, \hat{e}_n$ sind Schätzer für die nicht beobachtbaren Fehlervariablen $e_1, \ldots, e_n$. In diesem Abschnitt werden wir einige Resultate über das asymptotische Verhalten der Residuen angeben. Wir beginnen mit einigen Aussagen über die X_i. Die folgenden Ergebnisse sind elementar und wohlbekannt. Jedoch sind sie weit über die Literatur verstreut, weshalb wir im Folgenden auf Quellenangaben verzichten und statt dessen die elementaren Beweise vollständig ausführen. Gleiches gilt auch für Lemma 3.3.

Lemma 3.2

Unter der Voraussetzung (2.10) gilt:

$$\max_{1\le i\le n} |X_{i-1}| = \begin{cases} o_{\mathrm{P}}(n^{1/2}) & \text{für } |\rho| < 1 \\ O_{\mathrm{P}}(n^{1/2}) & \text{für } |\rho| = 1 \\ O_{\mathrm{P}}(|\rho|^n) & \text{für } |\rho| > 1 \end{cases} \tag{3.3}$$

$$\sum_{i=1}^n |X_{i-1}| = \begin{cases} O_{\mathrm{P}}(n) & \text{für } |\rho| < 1 \\ O_{\mathrm{P}}(n^{3/2}) & \text{für } |\rho| = 1 \\ O_{\mathrm{P}}(|\rho|^n) & \text{für } |\rho| > 1 \end{cases} \tag{3.4}$$

$$\sum_{i=1}^n X_{i-1} = \begin{cases} O_{\mathrm{P}}(n^{1/2}) & \text{für } |\rho| < 1 \\ O_{\mathrm{P}}(n^{3/2}) & \text{für } \rho = 1 \\ O_{\mathrm{P}}(n^{1/2}) & \text{für } \rho = -1 \\ O_{\mathrm{P}}(|\rho|^n) & \text{für } |\rho| > 1 \end{cases} \tag{3.5}$$

$$\sum_{i=1}^n X_{i-1}e_i = \begin{cases} O_{\mathrm{P}}(n^{1/2}) & \text{für } |\rho| < 1 \\ O_{\mathrm{P}}(n) & \text{für } |\rho| = 1 \\ O_{\mathrm{P}}(|\rho|^n) & \text{für } |\rho| > 1 \end{cases} \tag{3.6}$$

$$\sum_{i=1}^n X_{i-1}^2 = \begin{cases} O_{\mathrm{P}}(n) & \text{für } |\rho| < 1 \\ O_{\mathrm{P}}(n^2) & \text{für } |\rho| = 1 \\ O_{\mathrm{P}}(|\rho|^{2n}) & \text{für } |\rho| > 1. \end{cases} \tag{3.7}$$

Beweis Zu (3.3):

Ist $|\rho| < 1$, so gilt

$$\max_{1\le i\le n} |X_{i-1}| \le \max_{1\le i\le n} |e_i| \cdot \frac{1}{1-|\rho|}$$

und für $\varepsilon > 0$ gilt

$$\mathrm{P}\left(\max_{1\le i\le n} |e_i| \ge \sqrt{n}\varepsilon\right) \le n\mathrm{P}\left(e_1^2 \ge n\varepsilon^2\right) \le \frac{1}{\varepsilon^2}\mathrm{E}\left(e_1^2 1_{\{e_1^2 \ge n\varepsilon^2\}}\right) \longrightarrow 0 \qquad \text{für } n \to \infty$$

auf Grund der quadratischen Integrierbarkeit von e_1.

Ist $|\rho| = 1$, so ist $X_i = \sum_{j=1}^{i} (\pm 1)^{i-j} e_j$, und für jedes $C > 0$ gilt die Beziehung

$$\limsup_{n\to\infty} \mathrm{P}\left(\max_{1\le i\le n} |X_{i-1}| \ge \sqrt{n}C \right) \le \frac{\sigma^2}{C^2},$$

da wegen (2.10) die Kolmogorov-Ungleichung anwendbar ist.

Ist schließlich $|\rho| > 1$, so ist

$$\max_{1\le i\le n} |X_{i-1}| \le \sum_{i=1}^{n} |X_{i-1}|$$

und die Behauptung folgt aus (3.4).

Dies schließt den Beweis von (3.3) ab.

Zu (3.4):

Im Fall $|\rho| < 1$ gilt

$$\mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} |X_{i-1}| \right) \le \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=1}^{i} |\rho|^{i-j} \mathrm{E}(|e_i|) \le \frac{\mathrm{E}(|e_1|)}{1-|\rho|},$$

und dies impliziert die Behauptung mit der Markov-Ungleichung.

Der Fall $|\rho| = 1$ ergibt sich einfach mit (3.3) aus der Ungleichung

$$\sum_{i=1}^{n} |X_{i-1}| \le n \cdot \max_{1\le i\le n} |X_{i-1}|.$$

Schließlich gilt für $|\rho| > 1$ die Abschätzung

$$\mathrm{E}\left(\sum_{i=1}^{n} |X_{i-1}| \right) \le \sum_{i=1}^{n-1} \sum_{j=1}^{i} |\rho|^{i-j} \mathrm{E}(|e_i|) \le \frac{\mathrm{E}(|e_1|)}{(|\rho|-1)^2} |\rho|^n = O(|\rho|^n),$$

woraus die Behauptung mit der Markov-Ungleichung folgt.

Somit ist (3.4) bewiesen.

Zu (3.5):

Wir betrachten wieder zuerst den Fall $|\rho| < 1$. Für jedes $C > 0$ gilt

$$\begin{aligned} \mathrm{P}\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} \right| \ge C \right) &\le \frac{1}{nC^2} \mathrm{Var}\left(\sum_{i=1}^{n-1} X_i \right) \\ &= \frac{1}{nC^2} \mathrm{Var}\left(\sum_{j=1}^{n-1} \left(\sum_{i=j}^{n-1} \rho^{i-j} \right) e_j \right) \\ &\le \frac{\sigma^2}{(1-|\rho|)^2} \cdot \frac{1}{C^2}. \end{aligned}$$

Die rechte Seite der Ungleichungskette ist aber von n unabhängig und konvergiert für $C \to \infty$ gegen Null, was die Behauptung für diesen Fall beweist.

Im Fall $\rho = -1$ gilt

$$\sum_{i=1}^{n} X_{i-1} = \sum_{j=1}^{n-1} \left(\sum_{i=0}^{n-j-1} (-1)^i \right) e_j,$$

und wegen $\sum_{i=0}^{n-j-1} (-1)^i \in \{0, 1\}$ gilt für $C > 0$ die Ungleichung

$$\limsup_{n\to\infty} \mathrm{P} \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} \right| \geq C \right) \leq \frac{\sigma^2}{C^2},$$

was wie gewünscht $\sum_{i=1}^{n} X_{i-1} = O_{\mathrm{P}}(n^{1/2})$ impliziert.

Für $\rho = 1$ und $|\rho| > 1$ folgt die Behauptung sofort aus (3.4).

Damit sind alle Teile der Behauptung (3.5) gezeigt.

Zu (3.6):

Ist $|\rho| < 1$ und $C > 0$, so gilt

$$\begin{aligned} \mathrm{P} \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} e_i \right| \geq C \right) &\leq \frac{1}{nC^2} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \mathrm{E} \left(X_i X_j e_{i+1} e_{j+1} \right) \\ &= \frac{\sigma^2}{nC^2} \sum_{i=1}^{n-1} E \left(X_i^2 \right) \\ &\leq \frac{\sigma^4}{1-\rho^2} \cdot \frac{1}{C^2}. \end{aligned}$$

Dabei gilt das mittlere Gleichheitszeichen, da für $i \neq j$ entweder e_{i+1} oder e_{j+1} unabhängig von allen anderen Faktoren und der Erwartungswert somit Null ist. Die rechte Seite der Ungleichungskette konvergiert aber für $C \to \infty$ gegen Null.

Sei nun $|\rho| = 1$. Ebenso wie im vorherigen Fall erhalten wir für beliebiges $C > 0$ die Ungleichung

$$\mathrm{P} \left(\left| \frac{1}{n} \sum_{i=1}^{n} X_{i-1} e_i \right| \geq C \right) \leq \frac{\sigma^2}{n^2 C^2} \sum_{i=1}^{n-1} E \left(X_i^2 \right) \leq \frac{\sigma^4}{C^2},$$

und der Grenzübergang $C \to \infty$ impliziert die Behauptung.

Auch im Fall $|\rho| > 1$ beginnen wir den Beweis für jedes $C > 0$ mit der Abschätzung

$$\mathrm{P} \left(\left| \rho^{-n} \sum_{i=1}^{n} X_{i-1} e_i \right| \geq C \right) \leq \frac{\sigma^2}{\rho^{2n} C^2} \sum_{i=1}^{n-1} E \left(X_i^2 \right)$$

$$
\begin{aligned}
&\leq \frac{\sigma^4}{\rho^{2n}(\rho^2-1)C^2}\sum_{i=1}^{n-1}(\rho^2)^i \\
&\leq \frac{\sigma^4}{(\rho^2-1)^2}\cdot\frac{1}{C^2}.
\end{aligned}
$$

Hieraus folgt die behauptete stochastische Ordnung.

Dies vervollständigt den Beweis von (3.6).

Zu (3.7):

Im Fall $|\rho| < 1$ ist

$$
\mathrm{E}\left(\sum_{i=1}^{n} X_{i-1}^2\right) = \sigma^2\sum_{i=1}^{n-1}\sum_{j=1}^{i}(\rho^2)^{i-j} \leq \frac{\sigma^2}{1-\rho^2}\cdot n,
$$

was mit Hilfe der Markov-Ungleichung die Behauptung impliziert.

Weiter gilt

$$
\sum_{i=1}^{n} X_{i-1}^2 \leq n\cdot\left(\max_{1\leq i\leq n}|X_{i-1}|\right)^2
$$

was im Fall $|\rho| = 1$ mit (3.3) die Behauptung ergibt.

Schließlich erhalten wir im Fall $|\rho| > 1$ wie oben

$$
\sum_{i=1}^{n} X_{i-1}^2 \leq \left(\sum_{i=1}^{n}|X_{i-1}|\right)^2
$$

und die Behauptung folgt mit (3.4).

Damit ist mit (3.7) auch die letzte Behauptung des Lemmas gezeigt. □

Dieses Lemma erlaubt uns, nun einige asymptotische Eigenschaften der Residuen festzustellen.

Lemma 3.3

Unter den Voraussetzungen (2.10) und (3.2) gelten für alle $\rho \in \mathbb{R}$ die Aussagen

$$
\max_{1\leq i\leq n}|\hat{e}_i| = o_{\mathrm{P}}(n^{1/2}), \tag{3.8}
$$

$$
\sum_{i=1}^{n}(\hat{e}_i - e_i) = o_{\mathrm{P}}(n^{1/2}), \tag{3.9}
$$

$$\sum_{i=1}^{n} \hat{e}_i = O_{\mathrm{P}}(n^{1/2}), \tag{3.10}$$

$$\sum_{i=1}^{n} \left(\hat{e}_i^2 - e_i^2\right) = O_{\mathrm{P}}(1) \tag{3.11}$$

und

$$\frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 \longrightarrow \sigma^2 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.} \tag{3.12}$$

Beweis Zu (3.8):

Dies folgt unmittelbar aus (3.2), (3.3) und der Ungleichung

$$\max_{1 \le i \le n} |\hat{e}_i| \le \max_{1 \le i \le n} |e_i| + |\hat{\rho}_n - \rho| \max_{1 \le i \le n} |X_{i-1}|.$$

Zu (3.9):

Hierzu verwenden wir ebenfalls (3.2), (3.5) und die Gleichheit

$$\sum_{i=1}^{n} (\hat{e}_i - e_i) = (\rho - \hat{\rho}_n) \sum_{i=1}^{n} X_{i-1}.$$

Zu (3.10):

Dies folgt unmittelbar aus (3.9).

Zu (3.11):

Es ist

$$\sum_{i=1}^{n} \left(\hat{e}_i^2 - e_i^2\right) = (\hat{\rho}_n - \rho)^2 \sum_{i=1}^{n} X_{i-1}^2 - 2\,(\hat{\rho}_n - \rho) \sum_{i=1}^{n} X_{i-1} e_i,$$

und die Behauptung folgt aus (3.2), (3.6) und (3.7).

Zu (3.12):

Die Konvergenz folgt aus (3.11) und dem starken Gesetz der großen Zahlen.

Damit sind alle Aussagen des Lemmas bewiesen. □

3.1.3 Der empirische Prozeß der Residuen mit geschätztem Parameter im AR(1)-Modell

Genau wie im Fall unabhängig und identisch verteilter Daten existiert auch im AR(1)-Modell ein kanonischer Schätzer für $\mathrm{F}(\cdot, \vartheta_0)$, nämlich die *empirische Verteilungsfunktion der Residuen*

$$\mathrm{F}_{n,res}(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{\hat{e}_i \leq x\}} = \frac{1}{n} \sum_{i=1}^{n} 1_{\{e_i \leq x + (\hat{\rho}_n - \rho) X_{i-1}\}} \qquad \text{für } x \in \mathbb{R}.$$

Obwohl die Residuen nicht mehr unabhängig und identisch verteilt sind, sind doch die Prozeesse $(n^{1/2}(\mathrm{F}_{n,res} - \mathrm{F}(\cdot, \vartheta_0)))_{n \in \mathbb{N}}$ und $(n^{1/2}(\mathrm{F}_n - \mathrm{F}(\cdot, \vartheta_0)))_{n \in \mathbb{N}}$ asymptotisch äquivalent, das heißt die Differenz konvergiert gleichmäßig stochastisch gegen Null, falls geeignete Regularitätsbedingungen erfüllt sind. Dabei unterscheiden sich die Beweise für den stabilen, den semistabilen und den explosiven Fall wesentlich. So hat zunächst Boldin (1982) und später unter schwächeren Voraussetzungen Koul (1992) den stabilen Fall untersucht. Den semistabilen Fall findet man bei Ling (1998), wobei sowohl Boldin (1982) und Koul (1992) als auch Ling (1998) das allgemeinere Modell autoregressiver Prozesse p-ter Ordnung mit $p \in \mathbb{N}$ betrachten. Der explosive Fall im AR(1)-Modell wurde von Koul und Levental (1989) untersucht. Andererseits sind die Voraussetzungen an die e_i beziehungsweise an $\mathrm{F}(\cdot, \vartheta_0)$ in allen drei Fällen im wesentlichen dieselben, so daß wir im Folgenden hinreichende Bedingungen angeben können, die für alle $\rho \in \mathbb{R}$ die gewünschten Aussagen implizieren. Unter diesen Bedingungen können wir die oben genannten Ergebnisse in einem Satz zusammenfassen.

Satz 3.4

Gelten die Bedingungen (2.10) und (3.2) sowie

$$\mathrm{F}(\cdot, \vartheta_0) \text{ hat eine gleichmäßig stetige, } \lambda - f\ddot{u} \text{ positive Dichte } \mathrm{f}(\cdot, \vartheta_0), \tag{3.13}$$

so gilt

$$\sup_{x \in \mathbb{R}} \left| \sqrt{n} \left(\mathrm{F}_{n,res}(x) - \mathrm{F}_n(x) \right) \right| = o_{\mathrm{P}}(1) \tag{3.14}$$

Beweis Der Fall $|\rho| < 1$ ist Bemerkung 7.2.3 in Koul (1992), der Fall $|\rho| > 1$ andererseits ist Korollar 1 in Koul und Levental (1989). Im Fall $|\rho| = 1$ liefert der Hauptsatz in Ling (1998) die Aussage

$$\sup_{x \in \mathbb{R}} \left| \mathrm{F}_{n,res}(x) - \mathrm{F}_n(x) - (\hat{\rho}_n - \rho) \frac{\mathrm{f}(x, \vartheta_0)}{n} \sum_{i=1}^{n} X_{i-1} \right| = o_{\mathrm{P}}(1/\sqrt{n}),$$

und wegen der Voraussetzung (3.2) und da $\mathrm{f}(\cdot, \vartheta_0)$ als gleichmäßig stetige Dichte beschränkt ist, genügt es

$$\sum_{i=1}^{n} X_{i-1} = O_{\mathrm{P}}(n^{3/2})$$

zu zeigen. Dies gilt aber nach (3.5), so daß auch im Fall $|\rho| = 1$ die Behauptung folgt. □

Als direkte Folgerung dieses Satzes ergibt sich mit dem Satz von Cramér-Slutzky für alle $\rho \in \mathbb{R}$ die Aussage

$$\sqrt{n}\,(\mathrm{F}_{n,res} - \mathrm{F}(\cdot, \vartheta_0)) \xrightarrow{\mathcal{L}} \mathrm{B}^0(\mathrm{F}(\cdot, \vartheta_0)) \qquad \text{in } \mathrm{D}[-\infty, \infty] \qquad \text{für } n \to \infty,$$

also ein Analogon zu (2.1).

Bis hierher haben wir uns im AR(1)-Modell auf die Diskussion der Schätzung der Verteilungsfunktion $\mathrm{F}(\cdot, \vartheta_0)$ mit einem festen Verteilungsparameter ϑ_0 beschränkt. Jetzt wollen wir uns wieder dem empirischen Prozeß mit geschätztem Parameter zuwenden, um z.B. auch auf die allgemeinere Verteilungshypothese (2.2) testen zu können. Dabei ist zu beachten, daß in unserem Modell nun zwei Parameter geschätzt werden müssen: der Verteilungsparameter ϑ_0 und der Autoregressionsparameter ρ. Zur Schätzung des Verteilungsparameters müssen wir natürlich ebenfalls einen Schätzer auf Basis von $X_0, \ldots, X_n$ beziehungsweise $\hat{e}_1, \ldots, \hat{e}_n$ verwenden, den wir im Folgenden mit $\hat{\vartheta}_{n,res} = \hat{\vartheta}_{n,res}(\hat{e}_1, \ldots, \hat{e}_n)$ bezeichnen. Ein geeigneter Schätzer für ϑ_0 sollte eine (2.6) entsprechende stochastische Entwicklung haben, das heißt es sollte

$$\hat{\vartheta}_{n,res} - \vartheta_0 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i, \vartheta_0) + R_n \tag{3.15}$$

mit der Funktion L aus (2.6) und $\|R_n\| = o_{\mathrm{P}}(n^{-1/2})$ gelten. Dabei ist zu beachten, daß das erste Argument der Funktion L in (3.15) der Fehler e_i und nicht das Residuum $\hat{e}_i$ ist. Das bedeutet, daß $\hat{\vartheta}_{n,res}$ weiterhin eine Entwicklung in eine normierte Summe unabhängig und identisch verteilter Zufallsvariablen und einen asymptotisch vernachlässigbaren Restterm besitzt.

Um zu sehen, daß es überhaupt sinnvolle Schätzer mit der Entwicklung (3.15) gibt, betrachten wir wieder die Klasse $\mathcal{F} = \{\mathcal{N}(0, \sigma^2) : \sigma^2 \in (0, \infty)\}$ aller zentrierten Normalverteilungen. Im Abschnitt 2.1.2 hatten wir bereits gesehen, daß im Falle unabhängig und identisch verteilter e_i der Maximum-Likelihood-Schätzer

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2$$

für die wahre Varianz σ_0^2 die Entwicklung

$$\hat{\sigma}_n^2 - \sigma_0^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2 - \sigma_0^2 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i, \sigma_0^2)$$

mit $\mathrm{L}(x, \sigma_0^2) = x^2 - \sigma_0^2$ besitzt. Im AR(1)-Fall ist

$$\hat{\sigma}^2_{n,res} = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2$$

der entsprechende Schätzer für σ_0^2. Er erfüllt (3.15), da

$$\frac{1}{n} \sum_{i=1}^{n} \left(\hat{e}_i^2 - e_i^2 \right) = o_{\mathrm{P}}(n^{-1/2})$$

nach (3.11) gilt. Andere konkrete parametrische Verteilungsklassen werden wir im Abschnitt 3.3 diskutieren.

Bevor wir den nächsten Satz formulieren, wollen wir zur besseren Übersicht noch einmal die von nun an ständig benötigten Voraussetzungen an die Verteilungsklasse $\mathcal{F}$ und die Schätzer $\hat{\vartheta}_{n,res}$ und $\hat{\rho}_n$ zusammenfassen.

Die Verteilungsfunktion $\mathrm{F}(\cdot, \vartheta_0)$ besitzt eine gleichmäßig stetige, $\lambda - f\ddot{u}$ positive Dichte $\mathrm{f}(\cdot, \vartheta_0)$. (3.13)

Es ist $\mathrm{E}(e_1) = 0$ und $0 < \sigma^2 = \mathrm{E}(e_1^2) < \infty$. (2.10)

Es ist ϑ_0 ein innerer Punkt des Parameterraums Θ, und es existiert eine stetige Funktion $\Delta(\cdot, \vartheta_0) : \mathbb{R} \longrightarrow \mathbb{R}^d$, so daß

$$\sup_{x \in \mathbb{R}} |\mathrm{F}(x, \vartheta_0 + h) - \mathrm{F}(x, \vartheta_0) - \Delta(x, \vartheta_0)^t h| = o(\|h\|) \qquad \text{für } \|h\| \to 0 \tag{2.5}$$

gilt.

Es existiert eine meßbare Funktion $\mathrm{L}(\cdot, \vartheta_0) : \mathbb{R} \longrightarrow \mathbb{R}^d$ mit Erwartungswertvektor $\mathrm{E}(\mathrm{L}(e_1, \vartheta_0)) = 0$, existierender Kovarianzmatrix $\Lambda(\vartheta_0)$ und

$$\hat{\vartheta}_{n,res} - \vartheta_0 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i, \vartheta_0) + R_n. \tag{3.15}$$

Dabei gilt $\|R_n\| = o_{\mathrm{P}}(n^{-1/2})$.

Es ist $\hat{\rho}_n$ ein Schätzer für ρ mit

$$\hat{\rho}_n - \rho = \begin{cases} O_{\mathrm{P}}(n^{-1/2}) & \text{für } |\rho| < 1 \\ o_{\mathrm{P}}(n^{-1}) & \text{für } |\rho| = 1 \\ O_{\mathrm{P}}(|\rho|^{-n}) & \text{für } |\rho| > 1. \end{cases} \tag{3.2}$$

Auch an dieser Stelle weisen wir noch einmal darauf hin, daß zwischen diesen Voraussetzungen modellabhängig Redundanzen bestehen können. Wir hatten bereits ein Beispiel genannt, in dem $0 < \mathrm{E}(\mathrm{L}^2(e_1, \vartheta_0)) < \infty$ auch $0 < \mathrm{E}(e_1^2) < \infty$ impliziert.

Wir erhalten nun den folgenden Satz:

Satz 3.5

Gelten die obigen Bedingungen, das heißt (3.13), (2.10), (2.5), (3.15) und (3.2), so gilt

$$\sqrt{n}\left(\mathrm{F}_{n,res} - \mathrm{F}(\cdot, \hat{\vartheta}_{n,res})\right) \xrightarrow{\mathcal{L}} \mathrm{Z} \qquad \text{in } \mathrm{D}[-\infty, \infty] \qquad \text{für } n \to \infty. \tag{3.16}$$

Dabei ist Z derselbe Prozeß wie in (2.8).

Beweis Der Beweis beruht auf einer stochastischen Entwicklung des Prozesses in (3.16). Wir führen an dieser Stelle die Details nicht aus, da die Argumente denen gleichen, die nötig sind, wenn man den EL-Schätzer der Residuen an Stelle von $\mathrm{F}_{n,res}$ verwendet. Diesen Fall werden wir in den nächsten Abschnitten ausführlich darstellen. □

3.1.4 Schätzung unter Ausnutzung der Zentriertheit

Wie schon in Kapitel 2 liegt es nun nahe sich die Frage zu stellen, ob man unter Verwendung des empirischen Likelihood-Prinzips auch im AR(1)-Modell zu einem asymptotisch besseren Schätzer für $\mathrm{F}(\cdot, \vartheta_0)$ als $\mathrm{F}_{n,res}$ kommen kann. Wir halten dazu nochmals fest, daß die Voraussetzung $\mathrm{E}(e_1) = 0$ im AR(1)-Modell eine natürliche Voraussetzung an das Modell ist. Trotzdem wird die Zusatzinformation durch den Schätzer $\mathrm{F}_{n,res}$ nicht im Sinne einer möglichst präzisen Schätzung von $\mathrm{F}(\cdot, \vartheta_0)$ ausgenutzt.

Deshalb definieren wir analog zum Fall unabhängig und identisch verteilter Zufallsvariablen den EL-Verteilungsschätzer

$$\begin{aligned} \mathrm{F}_{n,res,z}(x) &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + t_{n,res}\hat{e}_i} 1_{\{\hat{e}_i \le x\}} \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + t_{n,res}\hat{e}_i} 1_{\{e_i \le x + (\hat{\rho}_n - \rho)X_{i-1}\}} \qquad \text{für } x \in \mathbb{R}. \end{aligned}$$

Analog zum vorigen Kapitel ist auch hier unter der Voraussetzung

$$\min_{1 \le i \le n} \hat{e}_i < 0 < \max_{1 \le i \le n} \hat{e}_i$$

die Gleichung

$$\sum_{i=1}^{n} \frac{1}{1 + t_{n,res}\hat{e}_i}\hat{e}_i = 0$$

für $t_{n,res}$ im Intervall

$$\left(\left(\frac{1}{n}-1\right)\frac{1}{\max_{1\le i\le n}\hat{e}_i}\,,\left(\frac{1}{n}-1\right)\frac{1}{\min_{1\le i\le n}\hat{e}_i}\right)$$

eindeutig lösbar, so daß in diesem Fall wieder $t_{n,res}$ und damit $\mathrm{F}_{n,res,z}$ wohldefiniert ist. Wir zeigen nun, daß die Voraussetzung $\min_{1\le i\le n}\hat{e}_i < 0 < \max_{1\le i\le n}\hat{e}_i$ für große n mit gegen Eins konvergierender Wahrscheinlichkeit erfüllt ist.

Lemma 3.6

Gelten die Bedingungen (3.13), (2.10) und (3.2), so gilt

$$\mathrm{P}\left(0 \notin (\min_{1\le i\le n}\hat{e}_i, \max_{1\le i\le n}\hat{e}_i)\right) \longrightarrow 0 \qquad \text{für } n\to\infty. \tag{3.17}$$

Beweis Für den diskreten Schätzer $\mathrm{F}_{n,res}$ gilt

$$\mathrm{P}\left(0 \notin \left(\mathrm{F}_{n,res}^{-1}(1/n), \mathrm{F}_{n,res}^{-1}(1)\right)\right) \longrightarrow 0 \qquad \text{für } n\to\infty,$$

falls wir

$$\sup_{x\in\mathbb{R}} |\mathrm{F}_{n,res}(x) - \mathrm{F}(x,\vartheta_0)| = o_{\mathrm{P}}(1)$$

zeigen können. Auf Grund der Voraussetzungen gilt (3.14), das heißt $\mathrm{F}_{n,res}$ erfüllt diese Bedingung gerade auf Grund des Satzes von Glivenko-Cantelli. Andererseits ist

$$\mathrm{F}_{n,res}^{-1}(1/n) = \min_{1\le i\le n}\hat{e}_i \qquad \text{und} \qquad \mathrm{F}_{n,res}^{-1}(1) = \max_{1\le i\le n}\hat{e}_i,$$

und das ergibt die Behauptung. □

Wir werden also im Folgenden immer annehmen, daß $\min_{1\le i\le n}\hat{e}_i < 0 < \max_{1\le i\le n}\hat{e}_i$ erfüllt und damit $\mathrm{F}_{n,res,z}$ wohldefiniert ist.

In einem ersten Schritt zeigen wir nun, daß es eine zu (2.14) analoge stochastische Entwicklung für $(n^{1/2}(\mathrm{F}_{n,res,z} - \mathrm{F}(\cdot,\vartheta_0)))_{n\in\mathbb{N}}$ gibt.

Lemma 3.7

Unter den Voraussetzungen (3.13), (2.10) und (3.2) gilt

$$\begin{aligned}\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x,\vartheta_0) &= \frac{1}{n}\sum_{i=1}^{n}\left(1_{\{e_i\le x\}} - \mathrm{F}(x,\vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2}e_i\right)\\ &\quad + R_n(x) \qquad \text{für } x\in\mathbb{R}\end{aligned} \tag{3.18}$$

mit U gemäß (2.15) und

$$\sup_{x\in\mathbb{R}} \|R_n(x)\| = o_{\mathrm{P}}(n^{-1/2}).$$

Beweis Wir unterscheiden den stabilen, den semistabilen und den explosiven Fall. Im stabilen Fall ist dies gerade Lemma 3.6 aus Genz (2001). Der explosive Fall wurde in Fink (2002) behandelt. Die behauptete Aussage ist eine Folgerung aus Satz 3.6. Es bleibt also der Fall $|\rho| = 1$ zu untersuchen. Wir zeigen zunächst nacheinander die Hilfsaussagen

$$t_{n,res} = O_{\mathrm{P}}(n^{-1/2}) \tag{3.19}$$

und

$$t_{n,res} = \frac{1}{\sigma^2} \cdot \frac{1}{n} \sum_{i=1}^{n} e_i + o_{\mathrm{P}}(n^{-1/2}). \tag{3.20}$$

Zum Beweis von (3.19) gehen wir vor wie in Owen (1990). Es gilt nämlich

$$\begin{aligned}
0 &= \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{e}_i}{1 + t_{n,res}\hat{e}_i} \right| \\
&= \left| \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i - \frac{1}{n} \sum_{i=1}^{n} \frac{t_{n,res}\hat{e}_i^2}{1 + t_{n,res}\hat{e}_i} \right| \\
&\geq |t_{n,res}| \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{e}_i^2}{1 + t_{n,res}\hat{e}_i} - \left| \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \right| \\
&\geq \frac{|t_{n,res}|}{1 + |t_{n,res}| \max_{1 \leq i \leq n} |\hat{e}_i|} \cdot \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 - \left| \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \right|,
\end{aligned}$$

also

$$\frac{|t_{n,res}|}{1 + |t_{n,res}| \max_{1 \leq i \leq n} |\hat{e}_i|} \cdot \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 \leq \left| \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \right| = O_{\mathrm{P}}(n^{-1/2})$$

gemäß (3.10). Auflösen nach $|t_{n,res}|$ ergibt (3.19) unter Beachtung von (3.8) und (3.12).

Nun zeigen wir (3.20), indem wir für $y \neq -1$ die Gleichung $\frac{1}{1+y} = 1 - y + \frac{y^2}{1+y}$ verwenden und deshalb

$$\begin{aligned}
0 &= \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i - t_{n,res} \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 + t_{n,res}^2 \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{e}_i^3}{1 + t_{n,res}\hat{e}_i} \\
&= \frac{1}{n} \sum_{i=1}^{n} e_i - \frac{1}{n} \sum_{i=1}^{n} (\hat{e}_i - e_i) - t_{n,res} \cdot \sigma^2 - t_{n,res} \left(\frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 - \sigma^2 \right) \\
&\quad + t_{n,res}^2 \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{e}_i^3}{1 + t_{n,res}\hat{e}_i}
\end{aligned}$$

schreiben. Der zweite Summand hat wegen (3.9) die Ordnung $o_{\mathrm{P}}(n^{-1/2})$, der vierte wegen (3.12) und (3.19), so daß nur noch der letzte Summand abzuschätzen ist. Für diesen Restterm gilt aber

$$\left| t_{n,res}^2 \frac{1}{n} \sum_{i=1}^n \frac{\hat{e}_i^3}{1 + t_{n,res}\hat{e}_i} \right| \leq t_{n,res}^2 \cdot \max_{1 \leq i \leq n} \frac{1}{1 + t_{n,res}\hat{e}_i} \cdot \max_{1 \leq i \leq n} |\hat{e}_i| \cdot \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 = o_{\mathrm{P}}(n^{-1/2})$$

nach (3.19), (3.8) und (3.12). Es gilt also

$$0 = \frac{1}{n} \sum_{i=1}^n e_i - t_{n,res} \cdot \sigma^2 + o_{\mathrm{P}}(n^{-1/2}),$$

und Auflösen nach $t_{n,res}$ schließt den Beweis von (3.20) ab.

Nun kommen wir zum Beweis der eigentlichen Aussage des Lemmas. Zunächst halten wir fest, daß wie oben

$$\begin{aligned}
\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x, \vartheta_0) = {} & \frac{1}{n} \sum_{i=1}^n 1_{\{\hat{e}_i \leq x\}} - t_{n,res} \frac{1}{n} \sum_{i=1}^n \hat{e}_i 1_{\{\hat{e}_i \leq x\}} \\
& + t_{n,res}^2 \frac{1}{n} \sum_{i=1}^n \frac{\hat{e}_i^2}{1 + t_{n,res}\hat{e}_i} 1_{\{\hat{e}_i \leq x\}} - \mathrm{F}(x, \vartheta_0) \\
= {} & \frac{1}{n} \sum_{i=1}^n \left(1_{\{e_i \leq x\}} - \mathrm{F}(x, \vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2} e_i \right) \\
& + \frac{1}{n} \sum_{i=1}^n \left(1_{\{\hat{e}_i \leq x\}} - 1_{\{e_i \leq x\}} \right) - t_{n,res} \frac{1}{n} \sum_{i=1}^n (\hat{e}_i - e_i) 1_{\{\hat{e}_i \leq x\}} \\
& - t_{n,res} \left(\frac{1}{n} \sum_{i=1}^n e_i 1_{\{\hat{e}_i \leq x\}} - \mathrm{U}(x) \right) - \left(t_{n,res} - \frac{1}{\sigma^2} \frac{1}{n} \sum_{i=1}^n e_i \right) \mathrm{U}(x) \\
& + t_{n,res}^2 \frac{1}{n} \sum_{i=1}^n \frac{\hat{e}_i^2}{1 + t_{n,res}\hat{e}_i} 1_{\{\hat{e}_i \leq x\}} \\
= {} & \frac{1}{n} \sum_{i=1}^n \left(1_{\{e_i \leq x\}} - \mathrm{F}(x, \vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2} e_i \right) \\
& + R_{1n}(x) - R_{2n}(x) - R_{3n}(x) - R_{4n}(x) + R_{5n}(x)
\end{aligned}$$

gilt. Es genügt nun zu zeigen, daß alle Restterme gleichmäßig in $x \in \mathbb{R}$ die Ordnung $o_{\mathrm{P}}(n^{-1/2})$ haben.

Zunächst gilt

$$\sup_{x \in \mathbb{R}} |R_{1n}(x)| = \sup_{x \in \mathbb{R}} |\mathrm{F}_{n,res}(x) - \mathrm{F}_n(x)|,$$

was gemäß (3.14) die gewünschte Ordnung für R_{1n} impliziert.

Um dies für R_{2n} zu zeigen, reicht wegen (3.19) die Abschätzung

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n (\hat{e}_i - e_i) 1_{\{\hat{e}_i \leq x\}} \right| \leq |\hat{\rho}_n - \rho| \frac{1}{n} \sum_{i=1}^n |X_{i-1}|,$$

denn die rechte Seite der Ungleichung hat gemäß (3.2) und (3.4) sogar die Ordnung $o_P(n^{-1/2})$.

Für R_{4n} erhalten wir die Behauptung aufgrund von (3.20) und der Beschränktheit von U. Es gilt nämlich

$$\sup_{x\in\mathbb{R}} |\mathrm{U}(x)| \leq \mathrm{E}(|e_1|).$$

Den Restterm R_{5n} können wir wie folgt abschätzen:

$$\sup_{x\in\mathbb{R}} |R_{5n}(x)| \leq t_{n,res}^2 \cdot \max_{1\leq i\leq n} \frac{1}{1+t_{n,res}\hat{e}_i} \cdot \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2,$$

und wegen (3.19), (3.8) und (3.12) erhalten wir sogar die Ordnung $O_{\mathrm{P}}(n^{-1})$.

Also bleibt die Behauptung nur noch für R_{3n} zu zeigen, wobei wegen (3.19) die Aussage

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i 1_{\{\hat{e}_i\leq x\}} - \mathrm{U}(x) \right| = o_{\mathrm{P}}(1) \tag{3.21}$$

ausreicht.

Um dies zu zeigen, benötigen wir etwas Notation. Seien im Folgenden e_i^+ und e_i^- der Positiv- beziehungsweise Negativteil von e_i und ebenso

$$\mathrm{U}^+(x) = \mathrm{E}(e_1^+ 1_{\{e_1\leq x\}}) \qquad \text{und} \qquad \mathrm{U}^-(x) = \mathrm{E}(e_1^- 1_{\{e_1\leq x\}}) \qquad \text{für } x \in \mathbb{R}. \tag{3.22}$$

Wegen der Stetigkeit von $\mathrm{F}(\cdot, \vartheta_0)$ und der Integrierbarkeit von e_1 sind U^+ und U^- stetig, beschränkt und monoton wachsend in x, insbesondere sind U^+ und U^- gleichmäßig stetig, und es gilt $\mathrm{U} = \mathrm{U}^+ - \mathrm{U}^-$.

Wenn wir also $\varepsilon > 0$ vorgeben, so existiert ein $\delta > 0$, so daß

$$\sup_{x\in\mathbb{R}} |\mathrm{U}^+(x+\delta) - \mathrm{U}^+(x)| < \frac{\varepsilon}{2}$$

gilt.

Andererseits können wir aus Lemma 2.2 in Zhang (1997) ableiten, daß sogar

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{e_i\leq x\}} - \mathrm{U}^+(x) \right| \xrightarrow{fs} 0 \qquad \text{für } n \to \infty \tag{3.23}$$

richtig ist.

Also erhalten wir auf dem Ereignis $\{|\hat{\rho}_n - \rho| \max_{1\le i\le n} |X_{i-1}| \le \delta\}$ für jedes $x \in \mathbb{R}$ die Ungleichungen

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{\hat{e}_i \le x\}} - \mathrm{U}^+(x) \quad &\le \quad \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{e_i \le x+\delta\}} - \mathrm{U}^+(x+\delta) \right| \\ &\quad + \sup_{x\in\mathbb{R}} |\mathrm{U}^+(x+\delta) - \mathrm{U}^+(x)| \\ &\le \quad \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{e_i \le x\}} - \mathrm{U}^+(x) \right| + \frac{\varepsilon}{2}\end{aligned}$$

und analog

$$\mathrm{U}^+(x) - \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{\hat{e}_i \le x\}} \le \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{e_i \le x\}} - \mathrm{U}^+(x) \right| + \frac{\varepsilon}{2}.$$

Hieraus wiederum folgt

$$\begin{aligned}\mathrm{P}\left(\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{\hat{e}_i \le x\}} - \mathrm{U}^+(x) \right| \ge \varepsilon \right) \quad &\le \quad \mathrm{P}\left(|\hat{\rho}_n - \rho| \max_{1\le i\le n} |X_{i-1}| > \delta \right) \\ &\quad + \mathrm{P}\left(\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{e_i \le x\}} - \mathrm{U}^+(x) \right| \ge \frac{\varepsilon}{2} \right),\end{aligned}$$

und beide Wahrscheinlichkeiten auf der rechten Seite konvergieren wegen (3.2), (3.3) und (3.23) gegen Null.

Also gilt

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{\hat{e}_i \le x\}} - \mathrm{U}^+(x) \right| = o_{\mathrm{P}}(1)$$

und analog

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^- 1_{\{\hat{e}_i \le x\}} - \mathrm{U}^-(x) \right| = o_{\mathrm{P}}(1),$$

woraus wir (3.21) wegen

$$\begin{aligned}\sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i 1_{\{\hat{e}_i \le x\}} - \mathrm{U}(x) \right| \quad &\le \quad \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^+ 1_{\{\hat{e}_i \le x\}} - \mathrm{U}^+(x) \right| \\ &\quad + \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} e_i^- 1_{\{\hat{e}_i \le x\}} - \mathrm{U}^-(x) \right|\end{aligned}$$

schließen können. Insgesamt folgt $\sup_{x\in\mathbb{R}} |R_{3n}(x)| = o_{\mathrm{P}}(n^{-1/2})$, und das Lemma ist bewiesen. □

3.2 Der modifizierte empirische Prozeß mit geschätztem Parameter

Wie im Kapitel 2 ist es nun unser Ziel, auch im AR(1)-Modell einen funktionalen Grenzwertsatz für den modifizierten empirischen Prozeß mit geschätztem Parameter, also den Prozeß $(n^{1/2}(\mathrm{F}_{n,res,z} - \mathrm{F}(\cdot, \hat{\vartheta}_{n,res})))_{n\in\mathbb{N}}$ zu zeigen. Um dies zu erreichen, leiten wir auch für diesen Prozeß eine geeignete stochastische Entwicklung her. Diese ist Gegenstand des nächsten Lemmas.

Lemma 3.8

Es sei $\rho \in \mathbb{R}$ beliebig, und die Bedingungen (3.13), (2.10), (2.5), (3.15) und (3.2) seien erfüllt. Dann gilt

$$\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x, \hat{\vartheta}_{n,res}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Y}_i(x) + R_n(x) \qquad \text{für } x \in \mathbb{R} \tag{3.24}$$

mit

$$\sup_{x\in\mathbb{R}} |R_n(x)| = o_{\mathrm{P}}(n^{-1/2}).$$

Dabei sind die Y_i gemäß (2.17) durch

$$Y_i(x) = 1_{\{e_i \le x\}} - \mathrm{F}(x, \vartheta_0) - \frac{\mathrm{U}(x)}{\sigma^2} e_i - \Delta(x, \vartheta_0)^t \mathrm{L}\,(e_i, \vartheta_0)$$

definiert.

Beweis Wir stellen zuerst wieder für alle $x \in \mathbb{R}$ fest, daß wir

$$\begin{aligned} \mathrm{F}_{n,res,z}(x) - \mathrm{F}(x, \hat{\vartheta}_{n,res}) &= (\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x, \vartheta_0)) - (\mathrm{F}(x, \hat{\vartheta}_{n,res}) - \mathrm{F}(x, \vartheta_0)) \\ &= A_{1n}(x) - A_{2n}(x) \end{aligned}$$

schreiben können. Aufgrund der Bedingungen (3.13), (2.10) und (3.2) gilt für $A_{1n}(x)$ die Entwicklung (3.18), und es genügt folglich, $A_{2n}(x)$ zu untersuchen. Da $\hat{\vartheta}_{n,res}$ gemäß (3.15) insbesondere konsistent für ϑ_0 ist, liegt $\hat{\vartheta}_{n,res}$ für große n mit gegen Eins konvergierender Wahrscheinlichkeit in einer Umgebung von ϑ_0, so daß wie im Beweis von (2.16) die Entwicklung

$$A_{2n}(x) = \Delta(x, \vartheta_0)^t \frac{1}{n}\sum_{i=1}^{n} \mathrm{L}(e_i, \vartheta_0) + R_n'(x)) \tag{3.25}$$

mit

$$\sup_{x\in\mathbb{R}} |R_n'(x)| = o_{\mathrm{P}}(n^{-1/2})$$

aus (2.5) und (3.15) folgt. Insbesondere implizieren (2.5) und (3.15) die Aussage

$$\sup_{x\in\mathbb{R}} |\mathrm{F}(x,\hat{\vartheta}_{n,res}) - \mathrm{F}(x,\vartheta_0)| = o_{\mathrm{P}}(1). \tag{3.26}$$

Wenn man die Entwicklungen für A_{1n} und A_{2n} zusammenfügt, ist das Lemma gezeigt. □

Damit sind wir in der Lage, das Hauptresultat dieses Kapitels zu formulieren.

Satz 3.9

Unter den Voraussetzungen (3.13), (2.10), (2.5), (3.15) und (3.2) gilt für alle $\rho \in \mathbb{R}$ die Verteilungskonvergenzaussage

$$\sqrt{n}\left(\mathrm{F}_{n,res,z} - \mathrm{F}(\cdot,\hat{\vartheta}_{n,res})\right) \xrightarrow{\mathcal{L}} \mathrm{W} \qquad \text{in } \mathrm{D}[-\infty,\infty] \qquad \text{für } n\to\infty. \tag{3.27}$$

Dabei ist W der Prozeß aus (2.19), also ein zentrierter Gaußprozeß mit stetigen Pfaden und der Kovarianzfunktion (2.20).

Beweis Die stochastische Entwicklung (3.24) impliziert die asymptotische Äquivalenz der Prozesse $(n^{1/2}(\mathrm{F}_{n,res,z} - \mathrm{F}(\cdot,\hat{\vartheta}_{n,res})))_{n\in\mathbb{N}}$ und $(n^{-1/2}\sum_{i=1}^{n} Y_i)_{n\in\mathbb{N}}$. Die Verteilungskonvergenz von $(n^{-1/2}\sum_{i=1}^{n} Y_i)_{n\in\mathbb{N}}$ gegen W in $\mathrm{D}[-\infty,\infty]$ ist aber gerade der Beweis zu (2.19). □

Wir kommen nun noch einmal auf das eingangs formulierte Testproblem (2.2), also

$$\mathrm{H}_0 : \mathrm{F} \in \mathcal{F} \qquad \text{gegen} \qquad \mathrm{H}_1 : \mathrm{F} \notin \mathcal{F}$$

mit

$$\mathcal{F} = \left\{\mathrm{F}(\cdot,\vartheta) | \vartheta \in \Theta \subset \mathbb{R}^d\right\}$$

zurück.

Aus den Grenzwertsätzen (2.19) und (3.27) folgt

$$\sup_{x\in\mathbb{R}} \left|\sqrt{n}\left(\mathrm{F}_{n,z}(x) - \mathrm{F}(x,\hat{\vartheta}_n)\right)\right| \xrightarrow{\mathcal{L}} \sup_{x\in\mathbb{R}} |\mathrm{W}(x)| \qquad \text{für } n\to\infty$$

und

$$\sup_{x\in\mathbb{R}} \left|\sqrt{n}\left(\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x,\hat{\vartheta}_{n,res})\right)\right| \xrightarrow{\mathcal{L}} \sup_{x\in\mathbb{R}} |\mathrm{W}(x)| \qquad \text{für } n\to\infty$$

auf Grund der Stetigkeit der Supremumsnorm auf dem Raum $\mathrm{D}[-\infty,\infty]$.

Diese Verteilungkonvergenzaussage erlaubt es, kritische Werte für die Teststatistiken $\sup_{x\in\mathbb{R}} |\sqrt{n}(\mathrm{F}_{n,z}(x) - \mathrm{F}(x,\hat{\vartheta}_n))|$ und $\sup_{x\in\mathbb{R}} |\sqrt{n}(\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x,\hat{\vartheta}_{n,res}))|$

bei endlichem Stichprobenumfang durch entsprechende Quantile der Verteilung der Grenzvariablen zu approximieren. Da die Supremumsnorm eines jeden Gaußprozesses mit stetigen Pfaden stetig verteilt ist (vgl. Ylvisaker (1968)), gilt dies auch für $\sup_{x \in \mathbb{R}} |\mathrm{W}(x)|$. Wenn wir also mit H die Verteilungsfunktion der Grenzvariablen und mit H^{-1} die zugehörige Quantilfunktion bezeichnen, so gilt für jedes $\alpha \in (0,1)$ die Konvergenz

$$\mathrm{P}\left(\sup_{x \in \mathbb{R}} \left|\sqrt{n}\left(\mathrm{F}_{n,z}(x) - \mathrm{F}(x, \hat{\vartheta}_n)\right)\right| \geq \mathrm{H}^{-1}(1-\alpha)\right) \longrightarrow \alpha \qquad \text{für } n \to \infty$$

und Analoges für die Teststatistik auf der Basis der Residuen.

Ungünstigerweise hängt die Verteilung der Grenzvariablen $\sup_{x \in \mathbb{R}} |\mathrm{W}(x)|$ jedoch in nichttrivialer Weise von der unbekannten Verteilungsfunktion $\mathrm{F}(\cdot, \vartheta_0)$ ab. Es ist also im allgemeinen auch unter der Hypothese des Testproblems nicht mehr zu erwarten, daß $\sup_{x \in \mathbb{R}} |\mathrm{W}(x)|$ verteilungsfrei ist, was eine direkte Berechnung von Quantilen dieser Verteilung unmöglich macht. Eine Möglichkeit zur Konstruktion asymptotischer Tests für (2.2) ist das Bootstrap-Verfahren, das unter gewissen zusätzlichen Bedingungen die Berechnung asymptotischer kritischer Werte ermöglicht, wie wir im Kapitel 4 sehen werden.

3.3 Beispiele

Wir werden nun einige konkrete Klassen von Verteilungsfunktionen untersuchen und nachweisen, daß für sie die funktionalen Grenzwertsätze aus Kapitel 2 und 3 gelten. Als erstes Beispiel betrachten wir

$$\mathcal{F}_1 = \{\mathcal{N}(0, \sigma_0^2) : \sigma_0^2 \in (0, \infty)\},$$

das Skalenmodell der *Normalverteilung*. Setzen wir

$$\mathrm{f}(x) = \frac{1}{2}\exp(-|x|) \qquad \text{für } x \in \mathbb{R}$$

die Dichte der *Doppelexponentialverteilung*, so ist unser zweites Beispiel das zugehörige Skalenmodell

$$\mathcal{F}_2 = \left\{\frac{1}{\sigma}\mathrm{f}\left(\frac{x}{\sigma}\right) : \sigma \in (0, \infty)\right\}.$$

Das dritte Beispiel ist das Skalenmodell der t_β-*Verteilung*, also

$$\mathcal{F}_3 = \left\{\frac{1}{\sigma}\mathrm{t}_\beta\left(\frac{x}{\sigma}\right) : \sigma \in (0, \infty)\right\}$$

mit der zu einem beliebigen $\beta > 0$ definierten Dichte

$$\mathrm{t}_\beta(x) = C(\beta)\left(1 + \frac{x^2}{\beta}\right)^{-\frac{\beta+1}{2}} \qquad \text{für } x \in \mathbb{R}$$

und

$$C(\beta) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta}{2}\right)\Gamma\left(\frac{1}{2}\right)} \cdot \frac{1}{\sqrt{\beta}}.$$

Schließlich ist

$$\mathcal{F}_4 = \left\{\frac{1}{\sigma}\mathrm{f}\left(\frac{x}{\sigma}\right) : \sigma \in (0, \infty)\right\}$$

mit

$$\mathrm{f}(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2} \qquad \text{für } x \in \mathbb{R}$$

das Skalenmodell der *logistischen Verteilung* unsere letzte Beispielklasse.

Für diese Beispiele zeigen wir zunächst die Voraussetzungen (2.5), (2.10) und (3.13), die ja rein analytische Eigenschaften der Modellklassen sind. Dazu betrachten wir die folgende Kollektion hinreichender Bedingungen.

(R0) Die Funktion $\mathrm{f} : \mathbb{R} \longrightarrow (0, \infty)$ ist eine (Lebesgue-)Dichte,

(R1) f ist gleichmäßig stetig,

(R2) f ist symmetrisch bezüglich 0,

(R3) f ist auf $(0, \infty)$ stetig differenzierbar mit Ableitung f',

(R4)

$$\lim_{x \downarrow 0} x\mathrm{f}'(x) = 0,$$

(R5)

$$I_\mathrm{f} = \int_{-\infty}^{\infty} x^2 \frac{\mathrm{f}'(x)^2}{\mathrm{f}(x)} dx < \infty,$$

(R6)

$$\int_{-\infty}^{\infty} x^2 \mathrm{f}(x) dx < \infty.$$

Unter diesen Voraussetzungen formulieren und beweisen wir das folgende Lemma.

Lemma 3.10

Gegeben seien die Bedingungen (R0)-(R6). Dann gelten für die zu f gehörige Skalenfamilie für jedes $\sigma_0 \in (0, \infty)$ die Bedingungen (2.10), (3.13) und (2.5).

Beweis Offensichtlich gilt (2.10) wegen (R6) und (R2), und (3.13) folgt aus (R0) und (R1), so daß (2.5) zu zeigen bleibt. Wir bezeichnen im Folgenden mit Q die zu f gehörige Wahrscheinlichkeitsverteilung und mit F die zugehörige Verteilungsfunktion. Die Voraussetzung (R5) impliziert mit Hilfe der Cauchy-Schwarz-Ungleichung insbesondere

$$\int_{-\infty}^{\infty} |x\mathrm{f}'(x)|dx = \int_{\mathbb{R}} \left| x\frac{\mathrm{f}'(x)}{\mathrm{f}(x)} \right| Q(dx) \leq \left(\int_{\mathbb{R}} x^2 \left(\frac{\mathrm{f}'(x)}{\mathrm{f}(x)} \right)^2 Q(dx) \right)^{1/2} < \infty.$$

Weiter können wir wegen (R4) die Funktion $x \longmapsto x\mathrm{f}'(x)$ stetig mit dem Wert Null in Null fortsetzen. Somit ergibt partielle Integration unter Beachtung von (R2) die Aussage

$$0 \leq C = \lim_{x\to\infty} x\mathrm{f}(x) = \int_0^{\infty} u\mathrm{f}'(u)du + \frac{1}{2},$$

was beweist, daß der Grenzwert $\lim_{x\to\infty} x\mathrm{f}(x)$ existiert und größer oder gleich Null ist. Wäre aber $C > 0$ so würde für alle großen x die Ungleichung

$$\mathrm{f}(x) \geq \frac{C}{2x}$$

gelten, was der Integrierbarkeit von f widerspricht. Somit ist $C = 0$. Wegen der Symmetrie von f gilt also

$$\lim_{x\to\pm\infty} x\mathrm{f}(x) = 0 \tag{3.28}$$

und außerdem

$$\int_{-\infty}^{\infty} x\mathrm{f}'(x)dx = -1. \tag{3.29}$$

Zum Beweis von (2.5) ist für beliebiges $\sigma \in (0, \infty)$ und $|h| < \sigma$ nun

$$R(h) = \sup_{x\in\mathbb{R}} \left| \mathrm{F}(x, \sigma + h) - \mathrm{F}(x, \sigma) - h\frac{\partial}{\partial\sigma}\mathrm{F}(x, \sigma) \right| = o(|h|)$$

und die Stetigkeit von $\frac{\partial}{\partial\sigma}\mathrm{F}(x, \sigma)$ in x zu zeigen. Dabei ist $\mathrm{F}(\cdot, \sigma)$ die zu $\frac{1}{\sigma}\mathrm{f}(\cdot/\sigma)$ gehörige Verteilungsfunktion, also gilt

$$\mathrm{F}(x, \sigma) = \mathrm{F}\left(\frac{x}{\sigma}\right) \qquad \text{für } x \in \mathbb{R}, \sigma \in (0, \infty)$$

sowie

$$\frac{\partial}{\partial\sigma}\mathrm{F}(x, \sigma) = -\frac{x}{\sigma^2}\mathrm{f}\left(\frac{x}{\sigma}\right) \qquad \text{für } x \in \mathbb{R}, \sigma \in (0, \infty),$$

und wegen (R1) ist $\frac{\partial}{\partial\sigma}\mathrm{F}(x, \sigma)$ stetig. Weiter gilt

$$\begin{aligned} R(h) &= \left| \int_{\sigma}^{\sigma+h} -\frac{x}{s^2}\mathrm{f}\left(\frac{x}{s}\right) ds + \frac{x}{\sigma^2}\mathrm{f}\left(\frac{x}{\sigma}\right) h \right| \\ &\leq |h| \sup \left\{ \sup_{x\in\mathbb{R}} \left| \frac{x}{s^2}\mathrm{f}\left(\frac{x}{s}\right) - \frac{x}{\sigma^2}\mathrm{f}\left(\frac{x}{\sigma}\right) \right| : s \text{ zwischen } \sigma \text{ und } \sigma + h \right\}, \end{aligned}$$

und dies ist von der gewünschten Ordnung, wenn wir

$$\lim_{h\to 0}\sup_{x\in\mathbb{R}}\left|\frac{x}{(\sigma+h)^2}\mathrm{f}\left(\frac{x}{\sigma+h}\right)-\frac{x}{\sigma^2}\mathrm{f}\left(\frac{x}{\sigma}\right)\right|=0$$

zeigen können. Für alle $x\in\mathbb{R}$ und alle $|h|<\sigma$ gilt aber

$$\begin{aligned}\left|\frac{x}{(\sigma+h)^2}\mathrm{f}\left(\frac{x}{\sigma+h}\right)-\frac{x}{\sigma^2}\mathrm{f}\left(\frac{x}{\sigma}\right)\right| &\le \frac{1}{\sigma+h}\left|\frac{x}{\sigma+h}\right|\left|\mathrm{f}\left(\frac{x}{\sigma+h}\right)-\mathrm{f}\left(\frac{x}{\sigma}\right)\right| \\ &\quad+\left|\frac{1}{(\sigma+h)^2}-\frac{1}{\sigma^2}\right||x|\mathrm{f}\left(\frac{x}{\sigma}\right) \\ &= \frac{1}{\sigma+h}S_1(x,h)+S_2(x,h),\end{aligned}$$

und für den zweiten Summanden gilt

$$\sup_{x\in\mathbb{R}}S_2(x,h)=|h|\frac{2\sigma+h}{(\sigma+h)^2\sigma}\sup_{x\in\mathbb{R}}\left|\frac{x}{\sigma}\right|\mathrm{f}\left(\frac{x}{\sigma}\right)\longrightarrow 0\qquad\text{für } h\to 0,$$

auf Grund von (3.28) und (R1). Es genügt also, noch

$$\sup_{x\in\mathbb{R}}S_1(x,h)\longrightarrow 0\qquad\text{für } h\to 0$$

zu zeigen. Seien dazu $\varepsilon, C>0$ beliebig. Wegen der gleichmäßigen Stetigkeit von f existiert ein $\delta>0$, so daß

$$|\mathrm{f}(s)-\mathrm{f}(t)|\le\frac{\varepsilon\sigma}{2C}\qquad\text{für } |s-t|\le\delta$$

gilt. Damit können wir für alle $|h|\le\min\{\frac{\sigma}{2},\frac{\delta\sigma^2}{2C}\}$ die Abschätzung

$$\sup_{|x|\le C}S_1(x,h)\le\varepsilon$$

folgern. Wegen (3.28) können wir schließlich C so groß wählen, daß

$$\left|\frac{x}{\sigma+h}\right|\mathrm{f}\left(\frac{x}{\sigma+h}\right)\le\frac{\varepsilon}{3}\qquad\text{sowie}\qquad\left|\frac{x}{\sigma+h}\right|\mathrm{f}\left(\frac{x}{\sigma}\right)\le\frac{2\varepsilon}{3}$$

für alle $|h|\le\frac{\sigma}{2}$ und alle $|x|>C$ gilt. Also gilt auch

$$\sup_{|x|>C}S_1(x,h)\le\varepsilon$$

für alle kleinen h und die Behauptung (2.5) ist gezeigt. □

Dieses Lemma beweist (2.10), (3.13) und (2.5) in unseren Beispielen. Ist nämlich φ die Dichte der Standardnormalverteilung, so gelten offensichtlich die Bedingungen (R0)-(R4) und (R6). Es muß also nur noch (R5) nachgewiesen werden. Dazu verwenden wir

$$\varphi'(x)=-x\varphi(x)\qquad\text{für } x\in\mathbb{R},$$

und damit gilt

$$I_\varphi = \int_{-\infty}^{\infty} x^4 \varphi(x) dx = 3.$$

Bezeichnen wir mit f die Dichte der Doppelexponentialverteilung, so sind (R0)-(R4) und (R6) wieder offensichtlich erfüllt. Zu (R5) stellen wir fest, daß

$$\mathrm{f}'(x) = -\frac{1}{2}\exp(-x) = -\mathrm{f}(x) \qquad \text{für } x \in (0, \infty)$$

gilt, und aus Symmetriegründen erhalten wir

$$I_{\mathrm{f}} = \int_0^{\infty} x^2 \exp(-x) dx = 2.$$

In unserem dritten Beispiel sind die Bedingungen (R0)-(R3) wiederum trivial erfüllt. Dabei gilt für die Ableitung

$$t'_\beta(x) = -C(\beta)\frac{\beta+1}{\beta} x \left(1 + \frac{x^2}{\beta}\right)^{-\frac{\beta+3}{2}} = -(\beta+1)\frac{x}{x^2+\beta}\mathrm{t}_\beta(x) \qquad \text{für } x \in \mathbb{R},$$

und wir können (R4) sofort ablesen. Es bleiben (R5) und (R6) zu zeigen. Zu (R5) berechnen wir

$$I_{\mathrm{t}_\beta} = 2C(\beta)\frac{(\beta+1)^2}{\beta^2}\int_0^{\infty} x^4 \left(1 + \frac{x^2}{\beta}\right)^{-\frac{\beta+5}{2}} dx = 3\frac{\beta+1}{\beta+3},$$

wie man durch zweifache partielle Integration sieht. Die Voraussetzung (R6) ist genau dann erfüllt, wenn $\beta > 2$ gilt, was man am Tailverhalten von $x^2\mathrm{t}_\beta(x)$ ablesen kann. Für alle anderen Regularitätsbedingungen ist diese Einschränkung nicht nötig. Schließlich untersuchen wir die Gültigkeit der Voraussetzungen (R0)-(R6) im Beispiel der logistischen Verteilung. Dabei sind (R0)-(R4) wegen

$$\mathrm{f}(x) = \frac{1}{\exp(x) + 2 + \exp(-x)} \qquad \text{für } x \in \mathbb{R}$$

wieder leicht zu sehen. Weiter gilt

$$\mathrm{f}'(x) = \frac{\exp(-x)(\exp(-x) - 1)}{(1 + \exp(-x))^3} = \frac{\exp(-x) - 1}{1 + \exp(-x)}\mathrm{f}(x) \qquad \text{für } x \in \mathbb{R},$$

weshalb auch (R6) gilt. Längere Berechnungen ergeben schließlich

$$I_{\mathrm{f}} = 2\int_0^{\infty} \frac{x^2 \exp(-x)(\exp(-x) - 1)^2}{(1 + \exp(-x))^4} dx = \frac{4}{3}\left(1 + \frac{\pi^2}{12}\right),$$

und in allen vier Beispielen sind die Regularitätsbedingungen (R0)-(R6) gezeigt.

Um die Gültigkeit der funktionalen Grenzwertsätze (2.19) und (3.27) zu gewährleisten, müssen jetzt noch geeignete Schätzer für ρ und σ_0 gefunden werden, die die Voraussetzungen (2.6) beziehungsweise (3.2) und (3.15) erfüllen. Im Beispiel der zentrierten Normalverteilung ist der Maximum-Likelihood-Schätzer für σ_0^2 bei unabhängig und identisch verteilten Zufallsvariablen gerade

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2,$$

und wie bereits gesehen gilt (2.6) mit der Einflußfunktion

$$\mathrm{L}(x, \sigma_0^2) = x^2 - \sigma_0^2.$$

Damit sind alle Voraussetzungen für (2.19) gezeigt.

Im AR(1)-Modell können wir das Paar (ρ, σ_0^2) mit Hilfe des Maximum-Likelihoodprinzips schätzen und erhalten so

$$\hat{\rho}_n = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_{i-1}^2},$$

also den Kleinst-Quadrate-Schätzer, als Schätzer für ρ und

$$\hat{\sigma}_{n,res}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{\rho}_n X_{i-1})^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2$$

als Schätzer für σ_0^2. Wie schon erwähnt erfüllt der Kleinst-Quadrate-Schätzer aber im Fall $|\rho| = 1$ nicht die Voraussetzung (3.2), weshalb schon im Normalverteilungsmodell die gemeinsame Likelihoodschätzung keinen geeigneten Schätzer liefert. Andererseits ist jeder Schätzer für ρ, der nur die Bedingung (3.2) erfüllt, im Hinblick auf die Konvergenzaussage (3.27) gleichwertig, denn weder $\hat{\rho}_n$ noch ρ selbst haben einen Einfluß auf den Grenzprozeß. Daher sollte man in der Praxis im AR(1)-Modell stets einen beliebigen Schätzer für ρ mit den Konvergenzordnungen (3.2) verwenden (z.B. den modifizierten Kleinst-Quadrate-Schätzer) und dann zur Schätzung des Verteilungsparameters (hier σ_0^2) die so erhaltenen Residuen anstelle der Fehler in den Verteilungsschätzer aus dem Modell unabhängig und identisch verteilter Zufallsvariablen einsetzen. Der Schätzer $\hat{\rho}_n$ für den Autoregressionsparameter erfüllt dann stets (3.2), so daß lediglich die Entwicklung (3.15) zu zeigen bleibt. In unserem konkreten Beispiel der zentrierten Normalverteilung ergibt sich

$$\hat{\sigma}_{n,res}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2$$

als Schätzer für σ_0^2, wobei die $\hat{e}_i$ nun die mit einem beliebigen (3.2) erfüllenden Schätzer $\hat{\rho}_n$ gebildeten Residuen sind. Bereits vorher wurde gezeigt, daß der so gewonnene Schätzer (3.15) erfüllt, weshalb (3.27) gezeigt ist. Damit ist unser erstes Beispiel abgeschlossen.

Diese Überlegungen zeigen, daß die gemeinsame Likelihoodschätzung des Parameterpaars (ρ, ϑ_0) nicht immer geeignete Schätzer liefert. Wir sehen im Folgenden aber auch noch einen weiteren Grund, auf die Likelihoodschätzung von ρ zu verzichten.

In unserem zweiten Beispiel, der Doppelexponentialverteilung, kann man im Falle unabhängig und identisch verteilter Zufallsvariablen leicht ausrechnen, daß

$$\hat{\sigma}_n = \frac{1}{n} \sum_{i=1}^{n} |e_i|$$

der ML-Schätzer für den Skalenparameter σ_0 ist, und offensichtlich gilt (2.6) mit der Einflußfunktion

$$\mathrm{L}(x, \sigma_0) = |x| - \sigma_0,$$

wegen $\mathrm{E}(|e_1|) = \sigma_0$.

Im AR(1)-Modell hingegen ist die gemeinsame ML-Schätzung für (ρ, σ_0) praktisch nicht mehr handhabbar. Verwendet man aber einen beliebigen Schätzer $\hat{\rho}_n$ für ρ, der (3.2) erfüllt, so ergibt einfach

$$\hat{\sigma}_{n,res} = \frac{1}{n} \sum_{i=1}^{n} |\hat{e}_i|$$

einen Schätzer für σ_0, von dem nun noch die Entwicklung (3.15) zu zeigen ist. Dies ist gleichwertig zu

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (|\hat{e}_i| - |e_i|) = o_{\mathrm{P}}(1),$$

und im Fall $|\rho| \geq 1$ gilt

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (|\hat{e}_i| - |e_i|) \right| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\hat{e}_i - e_i| = |\hat{\rho}_n - \rho| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |X_{i-1}| = o_{\mathrm{P}}(1)$$

gemäß (3.2) und (3.4). Im Fall $|\rho| < 1$ müssen wir genauer abschätzen. Sei dazu $\varepsilon > 0$ beliebig und $\delta > 0$ so, daß

$$\mathrm{P}(|e_1| \leq \delta) \leq \frac{\varepsilon^2 (1 - |\rho|)}{\mathrm{E}(|e_1|)}$$

gilt. Dann können wir auf dem Ereignis $\{|\hat{\rho}_n - \rho| \max_{1 \leq i \leq n} |X_{i-1}| \leq \delta\}$ wie folgt abschätzen:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} (|\hat{e}_i| - |e_i|) &= \frac{1}{n} \sum_{i=1}^{n} (|\hat{e}_i| - |e_i|) \, 1_{\{|e_i| > \delta\}} + \frac{1}{n} \sum_{i=1}^{n} (|\hat{e}_i| - |e_i|) \, 1_{\{|e_i| \leq \delta\}} \\ &= \frac{1}{n} \sum_{i=1}^{n} (\rho - \hat{\rho}_n) X_{i-1} \left(1_{\{e_i > \delta\}} - 1_{\{e_i < -\delta\}} \right) \end{aligned}$$

$$\begin{aligned}
&+\frac{1}{n}\sum_{i=1}^{n}\left(|\hat{e}_i|-|e_i|\right)1_{\{|e_i|\le\delta\}}\\
=\;& A_n(\delta)+B_n(\delta),
\end{aligned}$$

und es genügt zu zeigen, daß beide Summanden die stochastische Ordnung $o_{\mathrm{P}}(n^{-1/2})$ haben. Bei der Behandlung $A_n(\delta)$ wiederum reicht es,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{i-1}\left(1_{\{e_i>\delta\}}-1_{\{e_i<-\delta\}}\right)=O_{\mathrm{P}}(1) \tag{3.30}$$

zu beweisen. Da es sich wegen der Symmetrie der Doppelexponentialverteilung und der Unabhängigkeit von X_{i-1} und e_i hierbei um eine Summe zentrierter Variablen handelt, erhalten wir für beliebiges $C>0$ unter Verwendung der Tschebyscheff-Ungleichung

$$\begin{aligned}
\mathrm{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{i-1}\left(1_{\{e_i>\delta\}}-1_{\{e_i<-\delta\}}\right)\right|\ge \mathrm{C}\right) &\le \frac{1}{nC^2}\sum_{i=1}^{n}\mathrm{E}(X_{i-1}^2)\\
&\le \frac{\mathrm{E}(e_1^2)}{(1-\rho^2)C^2},
\end{aligned}$$

was (3.30) impliziert. Hieraus können wir sogar $A_n(\delta)=O_{\mathrm{P}}(n^{-1})$ folgern, so daß wir nur noch $B_n(\delta)$ untersuchen müssen.

Wegen

$$|B_n(\delta)|\le\frac{1}{n}\sum_{i=1}^{n}|\hat{e}_i-e_i|1_{\{|e_i|\le\delta\}}=|\hat{\rho}_n-\rho|\frac{1}{n}\sum_{i=1}^{n}|X_{i-1}|1_{\{|e_i|\le\delta\}}$$

genügt es

$$\frac{1}{n}\sum_{i=1}^{n}|X_{i-1}|1_{\{|e_i|\le\delta\}}=o_{\mathrm{P}}(1) \tag{3.31}$$

zu zeigen, um die gewünschte Ordnung für $B_n(\delta)$ zu beweisen. Auf Grund der Wahl von δ erhalten wir unter Verwendung der Markov-Ungleichung aber die Abschätzung

$$\begin{aligned}
\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n}|X_{i-1}|1_{\{|e_i|\le\delta\}}\ge\varepsilon\right) &\le \frac{\mathrm{P}\left(|e_1|\le\delta\right)}{n\varepsilon}\sum_{i=1}^{n}\mathrm{E}(|X_{i-1}|)\\
&\le \frac{\mathrm{E}(|e_1|)}{\varepsilon(1-|\rho|)}\mathrm{P}\left(|e_1|\le\delta\right)\\
&\le \varepsilon,
\end{aligned}$$

was den Beweis von (3.31) abschließt.

Damit haben wir auf dem Ereignis $\{|\hat{\rho}_n - \rho| \max_{1\leq i\leq n} |X_{i-1}| \leq \delta\}$ die gewünschte Ordnung für $n^{-1}\sum_{i=1}^n(|\hat{e}_i| - |e_i|)$ gezeigt. Das Gegenereignis hat aber gegen Null konvergierende Wahrscheinlichkeit, so daß auch im Fall $|\rho| < 1$ die Ordnung

$$\frac{1}{n}\sum_{i=1}^{n}(|\hat{e}_i| - |e_i|) = o_{\mathrm{P}}(n^{-1/2})$$

gegeben ist. Dies vervollständigt den Beweis von (3.15), und somit gilt auch im Beispiel der Doppelexponentialverteilung der Grenzwertsatz (3.27).

Sowohl bei der Normal- als auch bei der Doppelexponentialverteilung überzeugt man sich leicht, daß

$$\mathrm{E}(e_1\mathrm{L}(e_1, \sigma_0) = 0$$

gilt, so daß in diesen beiden Beispielen die Verwendung des EL-Verfahrens zur nichtparametrischen Verteilungsschätzung wirklich vorteilhaft ist.

Alle über die Doppelexponentialverteilung gemachten Aussagen können mit einigem technischem Aufwand auch auf das Skalenmodell einer beliebigen *exponentiellen Potenzverteilung* verallgemeinert werden. Dabei ist die exponentielle Potenzverteilung zum Parameter $\beta > 0$ die Verteilung mit der Dichte

$$\mathrm{f}_\beta(x) = \frac{\beta}{2\Gamma\left(\frac{1}{\beta}\right)}\exp\left(-|x|^\beta\right) \qquad \text{für } x \in \mathbb{R}.$$

Nun untersuchen wir wieder die Skalenmodelle der t_β-Verteilung und der logistischen Verteilung. Gemäß obigen Ausführungen können wir einen beliebigen (3.2) erfüllenden Schätzer $\hat{\rho}_n$ für ρ betrachten, und uns somit im Weiteren auf die Schätzung von σ_0 beschränken. Wie man leicht nachrechnet, kann die den ML-Schätzer für σ_0 bestimmende Gleichung nicht mehr geschlossen nach σ aufgelöst werden, so daß es nötig ist, mit allgemeinen Eigenschaften von ML-Schätzern zu argumentieren, wenn man (2.6) und (3.15) für diesen zeigen will.

Deshalb geben wir im Folgenden hinreichende Bedingungen an, unter denen (2.6) und (3.15) für den ML-Schätzer für σ_0 gelten und zeigen in einem zweiten Schritt, daß die t_β-Verteilung und die logistische Verteilung diese Bedingungen erfüllen. Natürlich kennt die klassische ML-Theorie im Modell unabhängig und identisch verteilter Zufallsvariablen bereits viele der folgenden Resultate, wir stellen sie trotzdem in größerer Ausführlichkeit dar, da Beweisteile oder zumindest Beweisideen auch im AR(1)-Modell relevant sind.

Gegeben seien ab jetzt die folgenden Regularitätsbedingungen:

(R0) Die Funktion $\mathrm{f} : \mathbb{R} \longrightarrow (0, \infty)$ ist eine (Lebesgue-)Dichte, (R1) f ist gleichmäßig stetig,

(R2) f ist symmetrisch bezüglich 0,

(D1) f ist auf ganz $\mathbb{R}$ stetig differenzierbar,

(R5)
$$I_{\mathrm{f}} = \int_{-\infty}^{\infty} x^2 \frac{\mathrm{f}'(x)^2}{\mathrm{f}(x)} dx < \infty,$$

(R6)
$$\int_{-\infty}^{\infty} x^2 \mathrm{f}(x) dx < \infty,$$

(R7) die Abbildung
$$(0, \infty) \ni x \longmapsto \frac{x\mathrm{f}'(x)}{\mathrm{f}(x)}$$
ist streng monoton fallend,

(R8)
$$\mathrm{f}_\infty^* = \lim_{x \to \infty} \frac{x\mathrm{f}'(x)}{\mathrm{f}(x)} < -1,$$

(R9)
$$\sup_{x \in \mathbb{R}} \frac{|\mathrm{f}'(x)|}{\mathrm{f}(x)} < \infty.$$

Dabei impliziert (D1) insbesondere die Voraussetzungen (R3), (R4) und die Stetigkeit von f.

Bekanntermaßen ist zu einem Vektor $\underline{x}_n \in \mathbb{R}^n$ die Likelihoodfunktion L_n durch

$$\mathrm{L}_n(\underline{x}_n, \sigma) = \frac{1}{\sigma^n} \prod_{i=1}^{n} \mathrm{f}\left(\frac{x_i}{\sigma}\right) > 0 \qquad \text{für } \sigma \in (0, \infty)$$

und somit die log-Likelihoodfunktion l_n durch

$$\mathrm{l}_n(\underline{x}_n, \sigma) = -n \log(\sigma) + \sum_{i=1}^{n} \log \mathrm{f}\left(\frac{x_i}{\sigma}\right) \qquad \text{für } \sigma \in (0, \infty)$$

definiert. Die log-Likelihoodfunktion ist stetig differenzierbar nach σ mit der Ableitung

$$\frac{\partial}{\partial \sigma} \mathrm{l}_n(\underline{x}_n, \sigma) = -\frac{n}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^{n} \frac{\frac{x_i}{\sigma} \mathrm{f}'\left(\frac{x_i}{\sigma}\right)}{\mathrm{f}\left(\frac{x_i}{\sigma}\right)} \qquad \text{für } \sigma \in (0, \infty).$$

Wir bezeichnen im Weiteren

$$\hat{n}(\underline{x}_n) = |\{i \in \{1, \ldots, n\} : x_i = 0\}|$$

und

$$B_n(f) = \{\underline{x}_n \in \mathrm{I\!R}^n : \mathrm{f}^*_\infty \frac{n - \hat{n}(\underline{x}_n)}{n} < -1\}.$$

Dann gilt das folgende Lemma.

Lemma 3.11

Gegeben seien die Bedingungen (R0), (R2), (D1), (R7) und (R8). Dann hat für beliebiges $\underline{x}_n \in B_n(f)$ die Likelihoodgleichung

$$\frac{\partial}{\partial \sigma} l_n(\underline{x}_n, \sigma) = 0$$

genau eine Lösung.

Beweis Die Likelihoodgleichung ist äquivalent zu

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\frac{x_i}{\sigma} \mathrm{f}'\left(\frac{x_i}{\sigma}\right)}{\mathrm{f}\left(\frac{x_i}{\sigma}\right)} = -1,$$

und für $\underline{x}_n \in B_n(f)$ ist wegen (D1) die linke Seite dieser Gleichung stetig in σ und konvergiert für $\sigma \to \infty$ gegen Null. Andererseits ist

$$\lim_{\sigma \downarrow 0} \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{x_i}{\sigma} \mathrm{f}'\left(\frac{x_i}{\sigma}\right)}{\mathrm{f}\left(\frac{x_i}{\sigma}\right)} = \mathrm{f}^*_\infty \frac{n - \hat{n}(\underline{x}_n)}{n} < -1$$

auf Grund von (R2), (R8) und der Stetigkeit von f. Schließlich stellt (R7) die Eindeutigkeit der Lösung sicher. □

Die zu $\underline{x}_n \in B_n(f)$ eindeutig bestimmte Lösung der Likelihoodgleichung $\hat{\sigma}_n(\underline{x}_n)$ heißt Maximum-Likelihood-Schätzer (ML-Schätzer) für σ_0. Mit maßtheoretischen Standardargumenten zeigt man die Borel-Meßbarkeit von $\mathrm{B}_n(f)$ und $\underline{x}_n \longmapsto \hat{\sigma}_n(\underline{x}_n)$.

Im Folgenden sei $\underline{e}_n = (e_1, \ldots, e_n)$ und $\hat{\underline{e}}_n = (\hat{e}_1, \ldots, \hat{e}_n)$. Dabei seien die e_i unabhängig und identisch nach Q_{σ_0} verteilt, wobei wir für $\sigma \in (0, \infty)$ mit Q_σ die zu $\frac{1}{\sigma}\mathrm{f}(\cdot/\sigma)$ gehörige Verteilung bezeichnen. Die $\hat{e}_i$ sind die mit Hilfe eines (3.2) erfüllenden Schätzers definierten Residuen. Da die e_i insbesondere stetig verteilt sind, gilt

$$\underline{e}_n \in B_n(f) \qquad \mathrm{P} - fs.$$

Für die zugehörigen Residuen gilt zumindest das folgende Lemma.

Lemma 3.12

Gegeben seien die Bedingungen (R0)-(R2), (D1) und (R6)-(R8). Dann gilt

$$\mathrm{P}\left(\underline{\hat{e}}_n \in B_n(f)\right) \longrightarrow 1 \qquad \text{für } n \to \infty.$$

Beweis Aus (R8) folgt die Existenz einer Zahl $0 < \delta < 1$ mit

$$\mathrm{f}_\infty^*(1-\delta) < -1,$$

weshalb

$$\mathrm{P}\left(\underline{\hat{e}}_n \notin B_n(f)\right) \leq \mathrm{P}\left(\frac{\hat{n}(\underline{\hat{e}}_n)}{n} > \delta\right)$$

gilt. Weiter implizieren (R0)-(R2) und (R6) die Aussagen (2.10) und (3.13). Somit gilt (3.14) und auf Grund des Satzes von Glivenko-Cantelli für die zu Q_{σ_0} gehörige Verteilungsfunktion $\mathrm{F}(\cdot, \sigma_0)$ auch

$$\sup_{x \in \mathbb{R}} \left|\mathrm{F}_{n,res}(x) - \mathrm{F}(x, \sigma_0)\right| = o_{\mathrm{P}}(1).$$

Hieraus folgt die Behauptung unter Beachtung von

$$\frac{\hat{n}(\underline{\hat{e}}_n)}{n} = \frac{1}{n} \sum_{i=1}^{n} 1_{\{\hat{e}_i = 0\}} = \mathrm{F}_{n,res}(0) - \mathrm{F}_{n,res}(0-).$$

□

Damit ist das Ereignis $\{\underline{\hat{e}}_n \notin B_n(f)\}$ für asymptotische Aussagen vernachlässigbar. Im Folgenden können wir also immer davon ausgehen, daß die ML-Schätzer $\hat{\sigma}_n = \hat{\sigma}_n(\underline{e}_n)$ auf Basis der Fehler und $\hat{\sigma}_{n,res} = \hat{\sigma}_n(\underline{\hat{e}}_n)$ auf Basis der Residuen wohldefiniert und eindeutig bestimmt sind.

Als nächstes zeigen wir die Konsistenz der ML-Schätzer $\hat{\sigma}_n$ und $\hat{\sigma}_{n,res}$. Im Modell unabhängig und identisch verteilter Zufallsvariablen ist der folgende Satz wohlbekannt; vgl. zum Beispiel Lehmann (1983).

Satz 3.13 *(vgl. Korollar 2.2 aus Abschnitt 6.2 in Lehmann (1983))*

Ist die Likelihoodgleichung

$$\frac{\partial}{\partial \sigma} l_n(\underline{e}_n, \sigma) = 0$$

fast sicher eindeutig lösbar, so ist der ML-Schätzer $\hat{\sigma}_n = \hat{\sigma}_n(\underline{e}_n)$ stark konsistent gegen den wahren Parameter σ_0.

Wie bereits gesehen stellen die Regularitätsbedingungen (R0)-(R2), (D1), (R5)-(R8) gerade die Gültigkeit dieses Satzes sicher. Ein wesentliches Element in der Beweisführung ist die *Kullback-Leibler-Information*, die für Wahrscheinlichkeitsmaße Q_1, Q_2 mit den zugehörigen Lebesgue-Dichten $\mathrm{q}_1, \mathrm{q}_2$ durch

$$K(Q_1, Q_2) = \int_{-\infty}^{\infty} \log\left(\frac{\mathrm{q}_1(x)}{\mathrm{q}_2(x)}\right) \mathrm{q}_1(x)dx$$

definiert ist. Bekanntermaßen gilt für die Kullback-Leibler-Information stets

$$0 \leq K(Q_1, Q_2) \leq \infty$$

und

$$K(Q_1, Q_2) = 0 \qquad \text{genau dann wenn} \qquad Q_1 = Q_2.$$

Außerdem definieren wir

$$\mathrm{l}_1(x, \sigma) = \log\left(\frac{1}{\sigma}\mathrm{f}\left(\frac{x}{\sigma}\right)\right) \qquad \text{für } x \in \mathbb{R}, \sigma \in (0, \infty).$$

Da uns über $\hat{\rho}_n$ und damit über $\hat{e}_1, \ldots, \hat{e}_n$ lediglich stochastische Ordnungsaussagen vorliegen, ist für $\hat{\sigma}_{n,res}$ keine starke Konsistenz gegen σ_0 zu erwarten. Es gilt aber das folgende stochastische Analogon des obigen Satzes.

Satz 3.14

Unter den Voraussetzungen (R0)-(R2), (D1) und (R5)-(R9) gilt für den ML-Schätzer im zugehörigen Skalenmodell

$$\hat{\sigma}_{n,res} = \hat{\sigma}_n(\underline{\hat{e}}_n) \longrightarrow \sigma_0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.} \tag{3.32}$$

Beweis Wie im klassischen Beweis für die starke Konsistenz des ML-Schätzers im Modell unabhängig und identisch verteilter Zufallsvariablen folgt (3.32) aus der Gültigkeit von

$$\frac{1}{n}\left(\mathrm{l}_n\left(\underline{\hat{e}}_n, \sigma_0\right) - \mathrm{l}_n\left(\underline{\hat{e}}_n, \sigma\right)\right) \longrightarrow K(Q_{\sigma_0}, Q_\sigma) \qquad \text{für } n \to \infty \tag{3.33}$$

P-stochastisch für beliebiges $\sigma \in (0, 2\sigma_0)$ mit den zu $\frac{1}{\sigma_0}\mathrm{f}(\cdot/\sigma_0)$ beziehungsweise $\frac{1}{\sigma}\mathrm{f}(\cdot/\sigma)$ gehörigen Wahrscheinlichkeitsverteilungen Q_{σ_0} und Q_σ. Zum Beweis von (3.33) schreiben wir

$$\frac{1}{n}\left(\mathrm{l}_n\left(\underline{\hat{e}}_n, \sigma_0\right) - \mathrm{l}_n\left(\underline{\hat{e}}_n, \sigma\right)\right) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{l}_1\left(\hat{e}_i, \sigma_0\right) - \mathrm{l}_1\left(e_i, \sigma_0\right)\right)$$

$$
\begin{aligned}
& +\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{l}_1\left(e_i,\sigma_0\right)-\mathrm{l}_1\left(e_i,\sigma\right)\right) \\
& +\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{l}_1\left(e_i,\sigma\right)-\mathrm{l}_1\left(\hat{e}_i,\sigma\right)\right) \\
= \; & S_{1n}+S_{2n}+S_{3n},
\end{aligned}
$$

und das starke Gesetz der großen Zahlen impliziert sofort die fast sichere Konvergenz von S_{2n} gegen die Kullback-Leibler-Information, so daß für jedes $0 < \sigma < 2\sigma_0$ noch

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{l}_1\left(e_i,\sigma\right)-\mathrm{l}_1\left(\hat{e}_i,\sigma\right)\right) = o_{\mathrm{P}}(1)$$

zu zeigen genügt. Der Mittelwertsatz impliziert

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{l}_1\left(\hat{e}_i,\sigma\right)-\mathrm{l}_1\left(e_i,\sigma\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{e}_i}{\sigma}-\frac{e_i}{\sigma}\right)\frac{\mathrm{f}'(\xi_{ni})}{\mathrm{f}(\xi_{ni})}$$

mit Zwischenstellen ξ_{ni} zwischen $\hat{e}_i/\sigma$ und e_i/σ, und hieraus wiederum folgt die stochastische Konvergenz von S_{1n} und S_{3n} gegen Null wegen

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{e}_i}{\sigma}-\frac{e_i}{\sigma}\right)\frac{\mathrm{f}'(\xi_{ni})}{\mathrm{f}(\xi_{ni})}\right| \leq \frac{1}{\sigma}\sup_{x\in\mathbb{R}}\frac{|\mathrm{f}'(x)|}{\mathrm{f}(x)}\frac{1}{n}\sum_{i=1}^{n}|\hat{e}_i-e_i|$$

unter Beachtung von (3.2) und (3.4) sowie (R9). Insgesamt folgt (3.33), und der Satz ist bewiesen. □

Somit sind unter geeigneten Regularitätsbedingungen sowohl $\hat{\sigma}_n$ als auch $\hat{\sigma}_{n,res}$ konsistente ML-Schätzer für σ_0. Wir weisen als nächstes die Voraussetzungen (D1) und (R7)-(R9) für die t_β-Verteilung und die logistische Verteilung nach. In beiden Fällen gilt (D1) offensichtlich. Weiter fällt $x\mathrm{t}'_\beta(x)/\mathrm{t}_\beta(x)$ monoton gegen

$$\lim_{x\to\infty}\frac{x\mathrm{t}'_\beta(x)}{\mathrm{t}_\beta(x)} = \lim_{x\to\infty}\frac{-(\beta+1)x^2}{x^2+\beta} = -(\beta+1) < -1,$$

was (R7) und (R8) impliziert. Außerdem rechnen wir leicht

$$\sup_{x\in\mathbb{R}}\frac{|\mathrm{t}'_\beta(x)|}{\mathrm{t}_\beta(x)} = \sup_{x\in\mathbb{R}}\frac{(\beta+1)|x|}{x^2+\beta} = \frac{(\beta+1)\sqrt{\beta}}{2\beta} < \infty$$

nach, was (R9) beweist. Damit sind alle bisher geforderten Voraussetzungen für jede t_β-Verteilung mit $\beta > 2$ erfüllt. Zu $(x_1,\ldots,x_n)\in B_n(\mathrm{t}_\beta)$ kann man die Likelihood-Gleichung

$$\sum_{i=1}^{n}\frac{x_i^2}{\beta\sigma^2+x_i^2} = \frac{n}{\beta+1}$$

zwar nicht explizit nach σ auflösen, dennoch kann natürlich numerisch die eindeutig bestimmte Lösung $\hat{\sigma}_n(\underline{x}_n)$ berechnet werden, und die so gebildeten ML-Schätzer $\hat{\sigma}_n = \hat{\sigma}_n(\underline{e}_n)$ und $\hat{\sigma}_{n,res} = \hat{\sigma}_n(\underline{\hat{e}}_n)$ sind auf Grund der bewiesenen Regularitätsbedingungen konsistent.

Im Beispiel der logistischen Verteilung gilt

$$\frac{x\mathrm{f}'(x)}{\mathrm{f}(x)} = \frac{x(\exp(-x)-1)}{1+\exp(-x)},$$

und dieser Quotient fällt für $x \to \infty$ monoton gegen $-\infty$, was (R7) und (R8) beweist. Schließlich ist (R9) offensichtlich wegen

$$\sup_{x\in\mathbb{R}} \frac{|\mathrm{f}'(x)|}{\mathrm{f}(x)} = \sup_{x\in\mathbb{R}} \frac{|\exp(-x)-1|}{1+\exp(-x)} = 1 < \infty$$

erfüllt. Die Likelihood-Gleichung

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\frac{x_i}{\sigma}\left(\exp\left(-\frac{x_i}{\sigma}\right)-1\right)}{1+\exp\left(-\frac{x_i}{\sigma}\right)} = -1$$

ist nicht geschlossen nach σ auflösbar, für numerisch bestimmte Lösungen gilt aber analog das oben Gesagte.

Wir kehren zur Betrachtung des allgemeinen Skalenmodells zurück. Für die unter (R0)-(R2), (D1) und (R5)-(R9) definierten ML-Schätzer $\hat{\sigma}_n$ und $\hat{\sigma}_{n,res}$ bleiben noch die Entwicklungen (2.6) und (3.15) zu zeigen. Es stellt sich heraus, daß

$$\mathrm{L}(x,\sigma_0) = \frac{1}{I(\sigma_0)}\frac{\partial}{\partial\sigma}\mathrm{l}_1(x,\sigma_0) = \frac{1}{I(\sigma_0)}\left(-\frac{1}{\sigma_0} - \frac{1}{\sigma_0}\frac{\frac{x}{\sigma_0}\mathrm{f}'\left(\frac{x}{\sigma_0}\right)}{\mathrm{f}\left(\frac{x}{\sigma_0}\right)}\right) \tag{3.34}$$

mit der *Fisher-Information*

$$I(\sigma) = \mathrm{E}\left(\left(\frac{\partial}{\partial\sigma}\log\frac{1}{\sigma}\mathrm{f}\left(\frac{e}{\sigma}\right)\right)^2\right) = \mathrm{E}\left(\left(\frac{\partial}{\partial\sigma}l_1(e,\sigma)\right)^2\right) \qquad \text{für } \sigma \in (0,\infty),$$

wobei e nach Q_σ verteilt ist, die (2.6) und (3.15) erfüllende Einflußfunktion ist. Dazu beweisen wir zunächst ein vorbereitendes Lemma.

Lemma 3.15

Unter den Bedingungen (R0), (R2), (D1) und (R5) gilt

$$I_{\mathrm{f}} > 1.$$

Beweis Wegen (3.29) gilt unter Beachtung der Cauchy-Schwarz-Ungleichung

$$1 = \left(\int_{-\infty}^{\infty} x \frac{\mathrm{f}'(x)}{\sqrt{\mathrm{f}(x)}} \sqrt{\mathrm{f}(x)} dx \right)^2 \leq I_{\mathrm{f}}$$

mit Gleichheit dann und nur dann, wenn

$$x\mathrm{f}'(x) = -\mathrm{f}(x) \qquad \lambda - f\ddot{u}$$

gilt. Dies widerspricht aber der Stetigkeit von $x\mathrm{f}'(x)$, da $\mathrm{f}(0) > 0$ ist. Damit ist das Lemma gezeigt. □

Nun können wir die in (2.6) und (3.15) geforderten analytischen Eigenschaften der Einflußfunktion und (2.24) für die Funktion L aus (3.34) zeigen.

Lemma 3.16

Unter den Voraussetzungen (R0)-(R2), (D1), (R5) und (R6) gelten für die gemäß (3.34) definierte Funktion L und eine nach Q_σ verteilte Zufallsvariable e die Aussagen

(i)
$$I(\sigma) = \frac{1}{\sigma^2} \left(I_{\mathrm{f}} - 1 \right), \tag{3.35}$$

(ii)
$$\mathrm{E}(\mathrm{L}(e, \sigma)) = 0, \tag{3.36}$$

(iii)
$$\mathrm{E}(\mathrm{L}^2(e, \sigma)) = \frac{1}{\mathrm{I}(\sigma)} \tag{3.37}$$

(iv) und
$$\mathrm{E}(e\mathrm{L}(e, \sigma)) = 0.$$

Beweis Es ist

$$I(\sigma) = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left(1 + \frac{\frac{x}{\sigma} \mathrm{f}' \left(\frac{x}{\sigma} \right)}{\mathrm{f} \left(\frac{x}{\sigma} \right)} \right)^2 \frac{1}{\sigma} \mathrm{f} \left(\frac{x}{\sigma} \right) dx = \frac{A}{\sigma^2}.$$

Substitution ergibt aber für das Integral

$$A = \int_{-\infty}^{\infty} \mathrm{f}(x) dx + 2 \int_{-\infty}^{\infty} x\mathrm{f}'(x) dx + I_{\mathrm{f}},$$

und daraus folgt (i) unter Beachtung von (3.29). Aus der Definition von L folgt hieraus trivial die Aussage (iii), und diese wiederum impliziert die Existenz des Erwartungswertes in (ii). Die Zentriertheit von L folgt wieder einfach aus (3.29), so daß nur noch (iv) zu zeigen bleibt. Aus

$$\mathrm{E}(|e\mathrm{L}(e,\sigma)|)^2 \leq \mathrm{E}(e^2)\mathrm{E}(\mathrm{L}^2(e,\sigma)) < \infty$$

folgt die Existenz des Erwartungswertes, wegen $\mathrm{E}(e) = 0$ gilt dann

$$\mathrm{E}(e\mathrm{L}(e,\sigma)) = -\frac{1}{\sigma I(\sigma)} \int_{-\infty}^{\infty} \frac{x^2}{\sigma^2}\mathrm{f}'\left(\frac{x}{\sigma}\right) dx = -\frac{1}{I(\sigma)} \int_{-\infty}^{\infty} x^2 \mathrm{f}'(x) dx,$$

und das letzte Integral ist auf Grund der Symmetrie von f gleich Null. Damit sind alle Aussagen des Lemmas gezeigt. □

Zum Beweis der Entwicklungen (2.6) und (3.15) setzen wir noch einmal weitere Regularitätsbedingungen fest. Seien im Folgenden

$$\mathrm{g}(x) = \frac{x\mathrm{f}'(x)}{\mathrm{f}(x)} \qquad \text{für } x \in \mathbb{R} \tag{3.38}$$

und

$$\mathrm{h}(x) = 1 + 2\mathrm{g}(x) + x^2 \frac{\mathrm{f}''(x)}{\mathrm{f}(x)} - \mathrm{g}^2(x) \qquad \text{für } x \in \mathbb{R} \tag{3.39}$$

definiert. Dann sind

(D2) f ist dreimal stetig differenzierbar,

(R10)

$$\int_{-\infty}^{\infty} x^2 |\mathrm{f}''(x)| dx < \infty,$$

(R11) es existiert eine Funktion $\mathrm{M}_{\sigma_0} : \mathbb{R} \longrightarrow (0, \infty)$ mit

$$\mathrm{E}(\mathrm{M}_{\sigma_0}(e_1)) < \infty$$

und ein $\varepsilon_0 > 0$ so, daß

$$\sup_{|\sigma-\sigma_0|\leq\varepsilon_0} \left| \frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(x,\sigma) \right| \leq \mathrm{M}_{\sigma_0}(x) \qquad \text{für } x \in \mathbb{R}$$

gilt,

(R12) h′ ist beschränkt,

(R13) g′ und g″ sind beschränkt

die zusätzlichen Regularitätsbedingungen. Natürlich impliziert (D2) insbesondere (D1).

Nun wäre es möglich, zum Beweis der Entwicklung (2.6) die Voraussetzungen des Satzes 2.7 nachzuweisen. Da wir aber Teile des Beweises ohnehin für den Beweis von (3.15) benötigen, führen wir ihn hier explizit aus. Grundlegend ist die Gleichung

$$0 = \sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \mathrm{l}_1(e_i, \sigma_n) = \sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \mathrm{l}_1(e_i, \sigma_0) + \sum_{i=1}^{n} \frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_i, \xi_{ni})(\hat{\sigma}_n - \sigma_0)$$

mit ξ_{ni} zwischen $\hat{\sigma}_n$ und σ_0 und dazu gleichwertig

$$\hat{\sigma}_n - \sigma_0 = \frac{-\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \mathrm{l}_1(e_i, \sigma_0)}{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_i, \xi_{ni})}.$$

Zum Beweis von (2.6) genügt es nun,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \mathrm{l}_1(e_i, \sigma_0) = O_{\mathrm{P}}(n^{-1/2}) \tag{3.40}$$

und

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_i, \xi_{ni}) \longrightarrow -I(\sigma_0) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch} \tag{3.41}$$

zu zeigen. Dabei gilt (3.40) offensichtlich wegen (3.36) und (3.37), und (3.41) können wir aus

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_i, \sigma_0) \longrightarrow -I(\sigma_0) \qquad \text{für } n \to \infty \qquad \mathrm{P} - fs \tag{3.42}$$

und

$$\frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_i, \xi_{ni}) - \frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_i, \sigma_0) \right) = o_{\mathrm{P}}(1) \tag{3.43}$$

ableiten. Die erste dieser beiden Aussagen folgt sofort aus dem starken Gesetz der großen Zahlen und dem folgenden Lemma.

Lemma 3.17

Unter den Voraussetzungen (R0)-(R2), (D2), (R5) und (R10) gilt

$$\mathrm{E}\left(\frac{\partial^2}{\partial\sigma^2} \mathrm{l}_1(e_1, \sigma_0) \right) = -I(\sigma_0).$$

Beweis Zunächst zeigen wir die Existenz des Erwartungswertes. Dazu stellen wir

$$\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(x,\sigma) = \frac{1}{\sigma^2}\mathrm{h}\left(\frac{x}{\sigma}\right) \qquad \text{für } x \in \mathbb{R}$$

und beliebiges $\sigma \in (0,\infty)$ fest, und die Integrale

$$\int_{-\infty}^{\infty} \mathrm{g}^2\left(\frac{x}{\sigma}\right)\frac{1}{\sigma}\mathrm{f}\left(\frac{x}{\sigma}\right)dx = \int_{-\infty}^{\infty} \mathrm{g}^2(x)\mathrm{f}(x)dx = I_{\mathrm{f}}$$

sowie

$$\int_{-\infty}^{\infty} \frac{x^2}{\sigma^2}\frac{\left|\mathrm{f}''\left(\frac{x}{\sigma}\right)\right|}{\mathrm{f}\left(\frac{x}{\sigma}\right)}\frac{1}{\sigma}\mathrm{f}\left(\frac{x}{\sigma}\right)dx = \int_{-\infty}^{\infty} x^2|\mathrm{f}''(x)|dx$$

sind gemäß (R5) und (R10) endlich. Außerdem gilt auf Grund von (R5) auch

$$\int_{-\infty}^{\infty} \left|\mathrm{g}\left(\frac{x}{\sigma}\right)\right|\frac{1}{\sigma}\mathrm{f}\left(\frac{x}{\sigma}\right)dx = \int_{-\infty}^{\infty} |x\mathrm{f}'(x)|dx < \infty$$

und

$$\int_{-\infty}^{\infty} x\mathrm{f}'(x)dx = -1, \tag{3.29}$$

wie wir bereits beim Beweis von Lemma 3.10 gezeigt haben. Analog zu (3.28) zeigt man

$$\lim_{x\to\pm\infty} x^2\mathrm{f}'(x) = 0,$$

und partielle Integration liefert unter Beachtung von (3.29) hiermit

$$\int_{-\infty}^{\infty} x^2\mathrm{f}''(x)dx = 2.$$

Insgesamt folgt

$$\mathrm{E}\left(\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_1,\sigma_0)\right) = \frac{1}{\sigma_0^2}\left(1 + 2\int_{-\infty}^{\infty} x\mathrm{f}'(x)dx + \int_{-\infty}^{\infty} x^2\mathrm{f}''(x)dx - I_{\mathrm{f}}\right) = -\frac{1}{\sigma_0^2}(I_{\mathrm{f}}-1),$$

und (3.35) impliziert die Behauptung. □

Damit gilt (3.42) und zum Beweis von (3.41) reicht es, (3.43) zu zeigen. Dies folgt aber aus dem nächsten Lemma. Die bewiesene Aussage ist etwas allgemeiner, da wir das Lemma in dieser Form auch zum Beweis von (3.15) verwenden können.

Lemma 3.18

Es gelten die Voraussetzungen (R0), (D2) und (R11). Sei weiter $(s_n)_{n\in\mathbb{N}}$ eine beliebige Folge von nichtnegativen Zufallsvariablen mit

$$s_n \longrightarrow \sigma_0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\xi_{ni})-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\sigma_0)\right)=o_{\mathrm{P}}(1) \tag{3.44}$$

für beliebige (zufällige) Zwischenstellen ξ_{ni} zwischen s_n und σ_0.

Beweis Sei $\varepsilon > 0$. Die Stetigkeit von $\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(x,\sigma)$ in σ, der Satz von Lebesgue und (R11) implizieren

$$\lim_{\delta\downarrow 0}\mathrm{E}\left(\sup_{|\sigma-\sigma_0|\le\delta}\left|\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_1,\sigma)-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_1,\sigma_0)\right|\right)=0,$$

was die Existenz einer Zahl $0<\delta_0\le\varepsilon_0$ mit

$$\mathrm{E}\left(\sup_{|\sigma-\sigma_0|\le\delta_0}\left|\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_1,\sigma)-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_1,\sigma_0)\right|\right)\le\varepsilon^2$$

sicherstellt. Dabei ist $\varepsilon_0 > 0$ gemäß Voraussetzung (R11) gegeben. Dann gilt

$$\begin{aligned}\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\xi_{ni})-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\sigma_0)\right)\right|\ge\varepsilon\right) &\le \mathrm{P}\left(|s_n-\sigma_0|>\delta_0\right)\\ &\quad+\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{|\sigma-\sigma_0|\le\delta_0}\left|\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\sigma)-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\sigma_0)\right|\ge\varepsilon\right)\\ &\le o(1)+\varepsilon\end{aligned}$$

auf Grund der Markov-Ungleichung, und das Lemma ist bewiesen. □

Damit gelten also (3.40) und (3.41), was die Entwicklung (2.6) mit der gemäß (3.34) definierten Einflußfunktion beweist. Es gilt also:

Satz 3.19

Unter den Voraussetzungen (R0)-(R2), (D2) und (R5)-(R11) ist der ML-Schätzer $\hat{\sigma}_n = \hat{\sigma}_n(\underline{e}_n)$ fast sicher eindeutig bestimmt, stark konsistent für σ_0 und hat die stochastische Entwicklung (2.6) mit der gemäß (3.34) definierten Einflußfunktion.

Es bleibt noch (3.15) zu zeigen. Die Vorgehensweise ist analog zu der im Modell unabhängig und identisch verteilter Zufallsvariablen. Grundlegend ist wieder die auf dem Mittelwertsatz beruhende Gleichung

$$\hat{\sigma}_{n,res}-\sigma_0=\frac{-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\sigma}\mathrm{l}_1(\hat{e}_i,\sigma_0)}{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(\hat{e}_i,\xi_{ni})}$$

mit Zwischenstellen ξ_{ni} zwischen $\hat{\sigma}_{n,res}$ und σ_0. Die behauptete Entwicklung folgt aus

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\sigma}\mathrm{l}_1(\hat{e}_i,\sigma_0)-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\sigma}\mathrm{l}_1(e_i,\sigma_0)=o_{\mathrm{P}}(n^{-1/2}) \tag{3.45}$$

und

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(\hat{e}_i,\xi_{ni})\longrightarrow -I(\sigma_0) \qquad \text{für } n\to\infty \qquad \text{P-stochastisch.} \tag{3.46}$$

Diese Aussagen sind Gegenstand der beiden nächsten Lemmata.

Lemma 3.20

Unter den Voraussetzungen (R0)-(R2), (D2) und (R5)-(R12) gilt

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(\hat{e}_i,\xi_{ni})\longrightarrow -I(\sigma_0) \qquad \text{für } n\to\infty \qquad \text{P-stochastisch.} \tag{3.46}$$

Beweis Es gilt

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(\hat{e}_i,\xi_{ni}) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\sigma_0)\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(\hat{e}_i,\xi_{ni})-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\xi_{ni})\right)\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\xi_{ni})-\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(e_i,\sigma_0)\right)\\
&= S_{1n}+S_{2n}+S_{3n},
\end{aligned}$$

und S_{1n} konvergiert gemäß (3.42) fast sicher gegen $-I(\sigma_0)$. Auf Grund von Lemma 3.18 und der Konsistenz von $\hat{\sigma}_{n,res}$ ist S_{3n} eine stochastische Nullfolge, so daß wir nur noch S_{2n} untersuchen müssen. Auf dem Ereignis $\{|\hat{\sigma}_{n,res}-\sigma_0|\leq\sigma_0/2\}$ gilt insbesondere für alle Zwischenstellen ξ_{ni} die Ungleichung

$$0<\frac{1}{\xi_{ni}}\leq\frac{2}{\sigma_0}.$$

Wegen (D2) ist die gemäß (3.39) definierte Funktion h differenzierbar, und der Mittelwertsatz und (R12) implizieren auf obigem Ereignis

$$\begin{aligned}
|S_{2n}| &= \left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\xi_{ni}^2}\left(\mathrm{h}\left(\frac{\hat{e}_i}{\xi_{ni}}\right)-\mathrm{h}\left(\frac{e_i}{\xi_{ni}}\right)\right)\right|\\
&\leq \frac{8}{\sigma_0^3}\sup_{x\in\mathbb{R}}|\mathrm{h}'(x)|\,\frac{1}{n}\sum_{i=1}^{n}|\hat{e}_i-e_i|.
\end{aligned}$$

Die rechte Seite der Ungleichung ist aber wegen (3.2) und (3.4) eine stochastische Nullfolge. Da andererseits

$$\mathrm{P}\left(|\hat{\sigma}_{n,res} - \sigma_0| > \frac{\sigma_0}{2}\right) \longrightarrow 0 \qquad \text{für } n \to \infty$$

gilt, ist das Lemma bewiesen. □

Lemma 3.21

Unter den Voraussetzungen (R0)-(R2), (D2) und (R5)-(R13) gilt

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \mathrm{l}_1(\hat{e}_i, \sigma_0) - \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\sigma} \mathrm{l}_1(e_i, \sigma_0) = o_{\mathrm{P}}(n^{-1/2}). \tag{3.45}$$

Beweis Zuerst untersuchen wir den Fall $|\rho| < 1$. Auf Grund von (D2) ist die gemäß (3.38) definierte Funktion g zweimal differenzierbar, und wegen

$$\frac{\partial}{\partial\sigma} \mathrm{l}_1(x, \sigma_0) = -\frac{1}{\sigma_0}\left(1 + \mathrm{g}\left(\frac{x}{\sigma}\right)\right) \qquad \text{für } x \in \mathbb{R}$$

gilt

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\sigma}\mathrm{l}_1(\hat{e}_i, \sigma_0) - \frac{\partial}{\partial\sigma}\mathrm{l}_1(e_i, \sigma_0)\right) = \frac{1}{\sigma_0}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathrm{g}\left(\frac{e_i}{\sigma_0}\right) - \mathrm{g}\left(\frac{\hat{e}_i}{\sigma_0}\right)\right) \\
&= \frac{1}{\sigma_0^2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{g}'\left(\frac{e_i}{\sigma_0}\right)(e_i - \hat{e}_i) + \frac{1}{\sigma_0^2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathrm{g}'(\xi_{ni}) - \mathrm{g}'\left(\frac{e_i}{\sigma_0}\right)\right)(e_i - \hat{e}_i) \\
&= S_{1n} + S_{2n}
\end{aligned}$$

mit Zwischenstellen ξ_{ni} zwischen $\hat{e}_i/\sigma_0$ und e_i/σ_o nach dem Mittelwertsatz. Mit S_{1n} beginnend zeigen wir, daß beide Summen stochastische Nullfolgen sind. Zunächst gilt

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{g}'\left(\frac{e_i}{\sigma_0}\right)(e_i - \hat{e}_i)\right| = |\hat{\rho}_n - \rho|\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{g}'\left(\frac{e_i}{\sigma_0}\right)X_{i-1}\right|,$$

und um die Diskussion von S_{1n} abzuschließen, genügt es wegen (3.2) für den zweiten Faktor auf der rechten Seite stochastische Beschränktheit nachzuweisen. Wegen (R13) ist aber g' bezüglich Q_{σ_0} integrierbar, und für jedes $C > 0$ gilt wegen der Zentriertheit der Fehler

$$\begin{aligned}
\mathrm{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathrm{g}'\left(\frac{e_i}{\sigma_0}\right)X_{i-1}\right| \geq C\right) &\leq \frac{1}{C^2}\left(\sup_{x\in\mathbb{R}}|\mathrm{g}'(x)|\right)^2\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(X_{i-1}^2\right) \\
&\leq \frac{1}{C^2}\left(\sup_{x\in\mathbb{R}}|\mathrm{g}'(x)|\right)^2\frac{\sigma_0^2}{1-\rho^2}
\end{aligned}$$

mithilfe der Tschebyscheff-Ungleichung. Die rechte Seite der Ungleichungskette konvergiert aber offensichtlich für $C \to \infty$ gegen Null.

Es bleibt S_{2n} zu betrachten. Hier gilt aber wegen der nochmaligen Differenzierbarkeit und der Beschränktheit der Ableitung

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathrm{g}'(\xi_{ni}) - \mathrm{g}'\left(\frac{e_i}{\sigma_0}\right)\right)(e_i - \hat{e}_i) \right| \leq \frac{1}{\sigma_0} \sup_{x \in \mathbb{R}} |\mathrm{g}''(x)| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{e}_i - e_i)^2$$

sowie

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{e}_i - e_i)^2 \leq \max_{1 \leq i \leq n} |\hat{e}_i - e_i| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\hat{e}_i - e_i| = o_{\mathrm{P}}(1)$$

gemäß (3.2), (3.4) und (3.3).

Ist $|\rho| \geq 1$, so gilt wegen (D2) und (R13) einfach

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\partial}{\partial \sigma} \mathrm{l}_1(\hat{e}_i, \sigma_0) - \frac{\partial}{\partial \sigma} \mathrm{l}_1(e_i, \sigma_0) \right) \right| \leq \sup_{x \in \mathbb{R}} |\mathrm{g}'(x)| \frac{1}{\sigma_0^2} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\hat{e}_i - e_i| = o_{\mathrm{P}}(1)$$

unter Beachtung von (3.2) und (3.4).

Damit ist (3.45) nachgewiesen. □

Die beiden Aussagen (3.45) und (3.46) implizieren nun die stochastische Entwicklung (3.15) und insgesamt gilt der folgende zu Satz 3.19 analoge Satz.

Satz 3.22

Unter den Voraussetzungen (R0)-(R2), (D2) und (R5)-(R13) ist der ML-Schätzer $\hat{\sigma}_{n,res} = \hat{\sigma}_n(\underline{\hat{e}}_n)$ mit gegen Eins konvergierender Wahrscheinlichkeit eindeutig bestimmt, konsistent für σ_0 und hat die stochastische Entwicklung (3.15) mit der gemäß (3.34) definierten Einflußfunktion.

Wir zeigen schließlich noch die Voraussetzungen (D2) und (R10)-(R13) in unserem dritten und vierten Beispiel. Im Falle der t_β-Verteilung ist (D2) offensichtlich erfüllt und die zweite Ableitung der Dichte ist

$$\mathrm{t}''_\beta(x) = \mathrm{t}_\beta(x) \left((\beta + 1) \frac{(\beta+2)x^2 - \beta}{(x^2 + \beta)^2} \right) \qquad \text{für } x \in \mathbb{R},$$

woraus wir auf Grund des Tailverhaltens sofort (R10) folgern können. Weiter ist

$$\mathrm{g}(x) = -(\beta + 1) \frac{x^2}{x^2 + \beta} \qquad \text{für } x \in \mathbb{R}$$

und

$$\mathrm{h}(x) = -\beta \cdot \frac{x^4 + (3\beta+1)x^2 - \beta}{(x^2+\beta)^2} \qquad \text{für } x \in \mathbb{R},$$

insbesondere ist h beschränkt durch eine Konstante C, so daß wir für $\sigma \geq \sigma_0/2$ die Abschätzung

$$\left| \frac{\partial^2}{\partial \sigma^2} \mathrm{l}_1(x,\sigma) \right| = \frac{1}{\sigma^2} \left| \mathrm{h}\left(\frac{x}{\sigma}\right) \right| \leq \frac{4C}{\sigma_0^2} \qquad \text{für } x \in \mathbb{R}$$

ableiten können. Die rechte Seite ist aber konstant also Q_{σ_0}-integrierbar, womit die Gültigkeit von (R11) gezeigt ist.

Wir berechnen nun die übrigen noch benötigten Ableitungen. Es ist

$$\mathrm{g}'(x) = -2\beta(\beta+1)\frac{x}{(x^2+\beta)^2} \qquad \text{für } x \in \mathbb{R},$$

$$\mathrm{g}''(x) = 2\beta(\beta+1)\frac{3x^2-\beta}{(x^2+\beta)^3} \qquad \text{für } x \in \mathbb{R}$$

und

$$\mathrm{h}'(x) = 2\beta(\beta+1)\frac{x^3-3\beta x}{(x^2+\beta)^3} \qquad \text{für } x \in \mathbb{R}.$$

Da diese Funktionen offensichtlich sämtlich beschränkt sind, gelten in unserem Beispiel der t_β-Verteilungen auch die Regularitätsbedingungen (R12) und (R13). Damit haben die ML-Schätzer $\hat{\sigma}_n$ und $\hat{\sigma}_{n,res}$ die Entwicklungen (2.6) beziehungsweise (3.15) mit der Einflußfunktion

$$\mathrm{L}(x,\sigma_0) = \frac{\sigma_0(\beta+3)}{2} \cdot \frac{x^2-\sigma_0^2}{x^2+\sigma_0^2\beta} \qquad \text{für } x \in \mathbb{R},$$

und es gelten die funktionalen Grenzwertsätze (2.19) und (3.27) sowie (2.24), was die Untersuchung des Beispiels abschließt.

Auch im Skalenmodell der logistischen Verteilung gilt natürlich (D2). Dabei ist

$$\mathrm{f}''(x) = \mathrm{f}(x)\frac{(2-\exp(-x))^2-3}{(1+\exp(-x))^2} \qquad \text{für } x \in \mathbb{R},$$

und auf Grund der Abschätzung

$$x^2|\mathrm{f}''(x)| \leq 2x^2\mathrm{f}(x) \qquad \text{für } x \in \mathbb{R}$$

sowie der Existenz zweiter Momente der logistischen Verteilung gilt (R10). Wir errechnen

$$\mathrm{g}(x) = \frac{x(\exp(-x)-1)}{1+\exp(-x)} \qquad \text{für } x \in \mathbb{R}$$

und

$$\begin{aligned}
\mathrm{h}(x) &= 1 - \frac{2x(x\exp(-x) - \exp(-2x) + 1)}{(1+\exp(-x))^2} \\
&= 1 + 2x(\exp(-x) - x - \exp(x))\mathrm{f}(x) \qquad \text{für } x \in \mathbb{R}
\end{aligned}$$

sowie

$$\lim_{x\to\infty} \frac{x\exp(-x) - \exp(-2x) + 1}{(1+\exp(-x))^2} = 1.$$

Also gilt

$$|\mathrm{h}(x)| \le 1 + C|x| \qquad \text{für } x \in \mathbb{R}$$

mit einer positiven Konstante C wegen der Symmetrie von h. Dies zeigt für $\sigma \geq \sigma_0/2$ die Abschätzung

$$\left|\frac{\partial^2}{\partial\sigma^2}\mathrm{l}_1(x,\sigma)\right| = \frac{1}{\sigma^2}\left|\mathrm{h}\left(\frac{x}{\sigma}\right)\right| \le \frac{4}{\sigma_0^2}\left(1 + \frac{2C}{\sigma_0}|x|\right) = \mathrm{M}_{\sigma_0}(x) \qquad \text{für } x \in \mathbb{R},$$

und M_{σ_0} ist wegen der Existenz erster Momente der logistischen Verteilung bezüglich Q_{σ_0} integrierbar, was (R11) beweist.

Wie im Beispiel der t_β-Verteilungen berechnen wir nun noch die folgenden Ableitungen. Es gilt

$$\mathrm{g}'(x) = \frac{\exp(-2x) - 2x\exp(-x) - 1}{(1+\exp(-x))^2} \qquad \text{für } x \in \mathbb{R},$$

$$\mathrm{g}''(x) = \frac{-2x\exp(-2x) + 2x\exp(-x) - 4\exp(-2x) - 4\exp(-x)}{(1+\exp(-x))^3} \qquad \text{für } x \in \mathbb{R}$$

sowie

$$\mathrm{h}'(x) = -2\mathrm{f}(x)\left(x\mathrm{g}(x) + 4x + \exp(x) - \exp(-x)\right) \qquad \text{für } x \in \mathbb{R}$$

und wegen der Existenz der Grenzwerte für $x \to \pm\infty$ sind alle diese Funktionen beschränkt, was (R12) und (R13) zeigt.

Damit besitzen die ML-Schätzer $\hat{\sigma}_n$ und $\hat{\sigma}_{n,res}$ auch hier die gewünschten Entwicklungen (2.6) beziehungsweise (3.15) mit der Einflußfunktion

$$\mathrm{L}(x,\sigma_0) = \frac{9}{3+\pi^2} \cdot \frac{x - \sigma_0 - (x+\sigma_0)\exp(-x/\sigma_0)}{1+\exp(-x/\sigma_0)} \qquad \text{für } x \in \mathbb{R}.$$

Somit gelten die funktionalen Grenzwertsätze (2.19) und (3.27) sowie die Bedingung (2.24) auch in unserem vierten Beispiel.

Kapitel 4

Bootstrapresultate

4.1 Modellvoraussetzungen und bekannte Ergebnisse

Das Bootstrap-Verfahren ist eine Möglichkeit, unbekannte Verteilungen, zum Beispiel die der Kolmogorov-Smirnov-Statistik für den Test (2.2), durch berechenbare Größen anzunähern. Im Falle unabhängig und identisch verteilter Daten, haben Stute et al. (1993) und später Babu und Rao (2004) ein Bootstrapresultat zu dem funktionalen Grenzwertsatz (2.8) bewiesen. Im Falle des AR(1)-Modells, aber auch insbesondere bei der Verwendung des EL-Prinzips zur Verteilungsschätzung in beiden Modellen stehen solche Bootstrap-Resultate noch aus. Wir beweisen in dieser Arbeit nur den Bootstrap im AR(1)-Modell und verzichten auf die Ausführung des (analogen) Beweises bei unabhängig und identisch verteilten Zufallsvariablen. Natürlich werden die dabei auftretenden Regularitätsvoraussetzungen durch Stute et al. (1993) und Babu und Rao (2004) inspiriert.

Wir benötigen zunächst etwas Notation. Sei $(\mathcal{A}_n)_{n\in\mathbb{N}}$ die Filtration, die durch

$$\mathcal{A}_n = \sigma(X_0,\ldots,X_n) = \sigma(e_1,\ldots,e_n) \qquad \text{für } n \geq 1$$

definiert wird. Weiter betrachten wir zu jeder beobachteten Folge $X_0,\ldots,X_n$ die *Bootstrap-Fehlervariablen* $e_{n1}^*,\ldots,e_{nn}^*$, die unter der bedingten Wahrscheinlichkeit $\mathrm{P}(\,\cdot\,|\mathcal{A}_n)$ unabhängig und identisch nach der (bekannten) Verteilung $\mathrm{F}(\cdot,\hat{\vartheta}_{n,res})$ verteilt seien. Hierzu muß die Verteilungsfunktion $\mathrm{F}(\cdot,\hat{\vartheta}_{n,res})$ natürlich definiert sein, das heißt $\hat{\vartheta}_{n,res}$ muß in Θ liegen. Da aber ϑ_0 ein innerer Punkt von Θ ist und aus (3.15) insbesondere

$$\hat{\vartheta}_{n,res} \longrightarrow \vartheta_0 \qquad \text{für } n\to\infty \qquad \text{P-stochastisch} \tag{4.1}$$

folgt, hat sogar für jede abgeschlossene Umgebung $V \subset \Theta$ von ϑ_0 das Ereignis $\{\hat{\vartheta}_{n,res} \notin V\}$ eine gegen Null konvergierende Wahrscheinlichkeit. Es ist damit für asymptotische Aussagen irrelevant, weshalb wir im Folgenden stets $\hat{\vartheta}_{n,res} \in \Theta$ voraussetzen. Wir schreiben für eine Folge von Zufallsvariablen $(Z_n^*)_{n\in\mathbb{N}}$ kurz

$$Z_n^* = O_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch}$$

beziehungsweise

$$Z_n^* = o_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch,}$$

falls für alle $\varepsilon, \varepsilon' > 0$ die Aussage

$$\lim_{C\to\infty} \limsup_{n\to\infty} \mathrm{P}\,(\mathrm{P}\,(|Z_n^*| \geq C|\mathcal{A}_n) \geq \varepsilon) = 0$$

beziehungsweise

$$\lim_{n\to\infty} \mathrm{P}\,(\mathrm{P}\,(|Z_n^*| \geq \varepsilon'|\mathcal{A}_n) \geq \varepsilon) = 0$$

gilt.

Obwohl die e_{ni}^* von n abhängen, wollen wir dies in der Notation künftig vernachlässigen und kurz e_i^* schreiben. Außerdem werden wir in Zukunft nicht zwischen einer Verteilungsfunktion $\mathrm{F}(\cdot, \vartheta)$ und der zugehörigen Wahrscheinlichkeitsverteilung unterscheiden und für beides stets $\mathrm{F}(\cdot, \vartheta)$ schreiben.

Zunächst stellen wir eine grundlegende Eigenschaft der e_i^* fest. Ist $\mathrm{h} : \mathbb{R} \longrightarrow \mathbb{R}$ eine meßbare und bezüglich $\mathrm{F}(\cdot, \vartheta_0)$ integrierbare Funktion, so gilt

$$\mathrm{E}\,(\mathrm{h}(e_1^*)|\mathcal{A}_n) \overset{fs}{=} \int_{\mathbb{R}} \mathrm{h}(x)\mathrm{F}(dx, \hat{\vartheta}_{n,res}),$$

denn $\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})$ ist ja gerade die bedingte Verteilung von e_1^* bei gegebenen $e_1, \ldots, e_n$. Im Folgenden identifizieren wir stets die fs-gleichen Versionen einer Zufallsvariablen miteinander und verzichten auf den Zusatz „fs“. Für jedes $x \in \mathbb{R}$ erhalten wir als Spezialfall obiger Gleichung

$$\mathrm{P}\,(e_1^* \leq x|\mathcal{A}_n) = \mathrm{F}(x, \hat{\vartheta}_{n,res}),$$

indem wir $\mathrm{h} = 1_{\{\cdot \leq x\}}$ setzen.

4.2 Der Bootstrap-Prozeß

4.2.1 Das asymptotische Verhalten der Bootstrap-Variablen

In diesem Abschnitt werden wir das asymptotische Verteilungsverhalten der e_i^* studieren. Um die gewünschten Ergebnisse zeigen zu können, müssen wir die folgende

Bedingung an die Verteilungsklasse $\mathcal{F}$ stellen:

Es existiert eine abgeschlossene Umgebung V von ϑ_0, so daß

$$\int_{\mathbb{R}} x\mathrm{F}(dx, \vartheta) = 0 \qquad \text{für } \vartheta \in V$$

und

$$\int_{\mathbb{R}} x^2\mathrm{F}(dx, \vartheta) \longrightarrow \int_{\mathbb{R}} x^2\mathrm{F}(dx, \vartheta_0) = \sigma^2 \qquad \text{für } \vartheta \to \vartheta_0 \tag{4.2}$$

gilt.

Diese Voraussetzung ist eine lokal gleichmäßige Version von (2.10). Aus den Voraussetzungen (4.2) und (2.5) erhalten wir für alle $x \in \mathbb{R}$ und alle $C > 0$ die folgenden Konvergenzaussagen:

$$\mathrm{F}(x, \vartheta) \longrightarrow \mathrm{F}(x, \vartheta_0) \qquad \text{für } \vartheta \to \vartheta_0, \tag{4.3}$$

$$\int_{\mathbb{R}} |y|\mathrm{F}(dy, \vartheta) \longrightarrow \int_{\mathbb{R}} |y|\mathrm{F}(dy, \vartheta_0) = \mathrm{E}(|e_1|) \qquad \text{für } \vartheta \to \vartheta_0, \tag{4.4}$$

$$\int_{\mathbb{R}} y^{\pm}1_{\{y \le x\}}\mathrm{F}(dy, \vartheta) \longrightarrow \int_{\mathbb{R}} y^{\pm}1_{\{y \le x\}}\mathrm{F}(dy, \vartheta_0) = \mathrm{U}^{\pm}(x) \qquad \text{für } \vartheta \to \vartheta_0 \tag{4.5}$$

und

$$\int_{\mathbb{R}} y^2 1_{\{|y| \ge C\}}\mathrm{F}(dy, \vartheta) \longrightarrow \int_{\mathbb{R}} y^2 1_{\{|y| \ge C\}}\mathrm{F}(dy, \vartheta_0) \qquad \text{für } \vartheta \to \vartheta_0. \tag{4.6}$$

Dabei ist (4.3) eine direkte Folgerung aus (2.5). Zum Beweis der Integralkonvergenzen verwendet man (4.3) und die Tatsache, daß (4.2) die gleichgradige Integrierbarkeit aller betrachteten Funktionen sicher stellt.

Da wir wegen (4.1), wie oben gesehen, stets $\hat{\vartheta}_{n,res} \in V$ annehmen können, gilt

$$\mathrm{E}\left(e_1^* | \mathcal{A}_n\right) = \int_{\mathbb{R}} x\mathrm{F}(dx, \hat{\vartheta}_{n,res}) = 0, \tag{4.7}$$

das heißt, die e_i^* sind bedingt zentriert. Ist weiter h einer der obigen Integranden, so bedeuten (4.1) und (4.3)-(4.6) gerade

$$\mathrm{E}\left(\mathrm{h}(e_1^*) | \mathcal{A}_n\right) \longrightarrow \mathrm{E}\left(\mathrm{h}(e_1)\right) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Einige für uns wichtige asymptotische Eigenschaften der e_i^* sind im folgenden Lemma zusammengefaßt.

Lemma 4.1

Gelten die Bedingungen (2.5), (4.2) und (3.15), so gilt

$$\sum_{i=1}^{n} e_i^* = O_{\mathrm{P}^*}(n^{1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch,} \tag{4.8}$$

$$\max_{1\le i\le n} |e_i^*| = o_{\mathrm{P}^*}(n^{1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch} \tag{4.9}$$

und

$$\frac{1}{n}\sum_{i=1}^{n} e_i^{*2} \longrightarrow \sigma^2 \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch.} \tag{4.10}$$

Beweis Zu (4.8):

Seien $\varepsilon > 0$ und $C > (2\sigma^2/\varepsilon)^{1/2}$. Dann gilt

$$\mathrm{P}\left(\mathrm{P}\left(\left|\sum_{i=1}^{n} e_i^*\right| \ge \sqrt{n}C\middle|\mathcal{A}_n\right) \ge \varepsilon\right) \le \mathrm{P}\left(\mathrm{E}\left(e_1^{*2}\middle|\mathcal{A}_n\right) \ge 2\sigma^2\right),$$

was wegen (4.1) und (4.2) bereits (4.8) beweist.

Zu (4.9):

Sind $\varepsilon, \varepsilon' > 0$, so gilt wie im Beweis von (3.3) die Ungleichung

$$\mathrm{P}\left(\max_{1\le i\le n} |e_i^*| \ge \sqrt{n}\varepsilon'\middle|\mathcal{A}_n\right) \le \frac{1}{\varepsilon'^2}\mathrm{E}\left(e_1^{*2}1_{\{e_1^{*2}\ge n\varepsilon'^2\}}\middle|\mathcal{A}_n\right).$$

Sei nun $C > 0$ so groß, daß

$$\mathrm{E}\left(e_1^2 1_{\{e_1^2>C\}}\right) \le \frac{\varepsilon}{2}$$

gilt. Ist n genügend groß, so folgt

$$\begin{aligned}\mathrm{P}\left(\mathrm{E}\left(e_1^{*2}1_{\{e_1^{*2}\ge n\varepsilon'^2\}}|\mathcal{A}_n\right) \ge \varepsilon\right) &\le \mathrm{P}\left(\mathrm{E}\left(e_1^{*2}1_{\{e_1^{*2}>C\}}|\mathcal{A}_n\right) \ge \varepsilon\right)\\ &\le \mathrm{P}\left(\left|\mathrm{E}\left(e_1^{*2}1_{\{e_1^{*2}>C\}}|\mathcal{A}_n\right) - \mathrm{E}\left(e_1^2 1_{\{e_1^2>C\}}\right)\right| \ge \frac{\varepsilon}{2}\right),\end{aligned}$$

und die letzte Wahrscheinlichkeit konvergiert auf Grund von (4.6) und (4.1) gegen Null, womit (4.9) bewiesen ist.

Zu (4.10):

Wählen wir ein beliebiges $\varepsilon' > 0$, so gilt für jedes $n \in \mathbb{N}$ und jedes $C > 0$ die Zerlegung

$$\begin{aligned}
&\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} e_i^{*2} - \sigma^2\right| \geq \varepsilon' \Big| \mathcal{A}_n\right) \\
\leq\ & \mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} e_i^{*2} 1_{\{|e_i^*| \leq C\}} - \mathrm{E}\left(e_1^{*2} 1_{\{|e_1^*| \leq C\}} | \mathcal{A}_n\right)\right| \geq \frac{\varepsilon'}{3} \Big| \mathcal{A}_n\right) \\
&+\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} e_i^{*2} 1_{\{|e_i^*| > C\}} - \mathrm{E}\left(e_1^{*2} 1_{\{|e_1^*| > C\}} | \mathcal{A}_n\right)\right| \geq \frac{\varepsilon'}{3} \Big| \mathcal{A}_n\right) \\
&+1_{\left\{\left|\mathrm{E}\left(e_1^{*2} | \mathcal{A}_n\right) - \sigma^2\right| \geq \frac{\varepsilon'}{3}\right\}} \\
=\ & P_{1n} + P_{2n} + P_{3n}.
\end{aligned}$$

Dabei konvergiert P_{3n} wegen (4.1) und (4.2) stochastisch gegen Null. Es bleiben also die beiden anderen Summanden zu diskutieren.

Weil die Summanden in P_{1n} durch C^2 beschränkt und insbesondere quadratisch integrierbar sind, folgt mit der Tschebyscheff-Ungleichung

$$P_{1n} \leq \frac{9C^4}{\varepsilon'^2 n},$$

und die rechte Seite der Ungleichung konvergiert gegen Null.

Zum Schluß bleibt P_{2n} zu untersuchen. Sei $\varepsilon > 0$ und $C > 0$ so groß, daß

$$\mathrm{E}\left(e_1^2 1_{\{|e_1| > C\}}\right) < \min\left\{\frac{\varepsilon'}{12}, \frac{\varepsilon\varepsilon'}{24}\right\}$$

ist, so erhalten wir

$$P_{2n} \leq \mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^{*2} 1_{\{|e_i^*| > C\}} \geq \frac{\varepsilon'}{12} \Big| \mathcal{A}_n\right) + 1_{\left\{\left|\mathrm{E}\left(e_1^{*2} 1_{\{|e_1^*| > C\}} | \mathcal{A}_n\right) - \mathrm{E}\left(e_1^2 1_{\{|e_1| > C\}}\right)\right| \geq \frac{\varepsilon'}{6}\right\}}.$$

Dabei ist der zweite Summand auf der rechten Seite der Ungleichung gemäß (4.1) und (4.6) eine stochastische Nullfolge. Andererseits gilt für den ersten

$$\begin{aligned}
&\mathrm{P}\left(\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^{*2} 1_{\{|e_i^*| > C\}} \geq \frac{\varepsilon'}{12} \Big| \mathcal{A}_n\right) \geq \varepsilon\right) \\
&\qquad \leq\ \mathrm{P}\left(\mathrm{E}\left(e_1^{*2} 1_{\{|e_1^*| > C\}} | \mathcal{A}_n\right) \geq \frac{\varepsilon\varepsilon'}{12}\right) \\
&\qquad \leq\ \mathrm{P}\left(\left|\mathrm{E}\left(e_1^{*2} 1_{\{|e_1^*| > C\}} | \mathcal{A}_n\right) - \mathrm{E}\left(e_1^2 1_{\{|e_1| > C\}}\right)\right| \geq \frac{\varepsilon\varepsilon'}{24}\right),
\end{aligned}$$

und der letzte Term verschwindet (4.1) und (4.6) folgend, wenn n gegen Unendlich geht. Insgesamt folgt die Gültigkeit von (4.10), was den Beweis des Lemmas abschließt. □

Es ist noch anzumerken, daß die Voraussetzungen für das gerade bewiesene Lemma noch abgeschwächt werden könnten. So ist zum Beispiel (3.15) im allgemeinen stärker als (4.1), und auch (2.5) wird nicht in voller Stärke benötigt. Da wir aber diese Voraussetzungen ohnehin für den funktionalen Grenzwertsatz (3.27) benötigen, den wir ja bootstrappen wollen, verzichten wir hier und im Folgenden darauf, schwächere hinreichende Bedingungen zu formulieren.

4.2.2 Der Bootstrap-Schätzer für die modifizierte empirische Verteilungsfunktion der Residuen

Um den Prozeß $(n^{1/2}(\mathrm{F}_{n,res,z} - \mathrm{F}(\cdot, \hat{\vartheta}_{n,res})))_{n\in\mathbb{N}}$ zu bootstrappen, ist es nötig, sowohl für $\mathrm{F}_{n,res,z}$ als auch für $\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})$ geeignete Bootstrap-Analoga auf Basis von $e_1^*, \ldots, e_n^*$ zu berechnen. Einen Bootstrap-Schätzer für $\mathrm{F}_{n,res,z}$ erhalten wir, indem wir den EL-Ansatz auf die Bootstrap-Variablen $e_1^*, \ldots, e_n^*$ anwenden. Dazu muß wie in den vorangegangenen Kapiteln das Ereignis $\{0 \notin (\min_{1\le i\le n} e_i^*, \max_{1\le i\le n} e_i^*)\}$ asymptotisch vernachlässigbar sein, was das folgende Lemma sicherstellt.

Lemma 4.2

Unter den Voraussetzungen (2.10), (2.5) und (3.15) gilt

$$\mathrm{P}\left(0 \notin (\min_{1\le i\le n} e_i^*, \max_{1\le i\le n} e_i^*) \Big| \mathcal{A}_n\right) \longrightarrow 0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Beweis Zunächst stellen wir fest, daß

$$\{0 \notin (\min_{1\le i\le n} e_i^*, \max_{1\le i\le n} e_i^*)\} \subseteq \bigcap_{i=1}^{n} \{e_i^* \le 0\} \cup \bigcap_{i=1}^{n} \{e_i^* \ge 0\}$$

gilt, und für jedes $\varepsilon > 0$ erhalten wir für alle großen n die Abschätzung

$$\begin{aligned} \mathrm{P}\left(\mathrm{P}\left(\bigcap_{i=1}^{n} \{e_i^* \le 0\} \Big| \mathcal{A}_n\right) \ge \varepsilon\right) &= \mathrm{P}\left(\left(\mathrm{F}(0, \hat{\vartheta}_{n,res}) - \mathrm{F}(0, \vartheta_0) + \mathrm{F}(0, \vartheta_0)\right)^n \ge \varepsilon\right) \\ &\le \mathrm{P}\left(\left|\mathrm{F}(0, \hat{\vartheta}_{n,res}) - \mathrm{F}(0, \vartheta_0)\right| > \frac{1 - \mathrm{F}(0, \vartheta_0)}{2}\right). \end{aligned}$$

Da $\mathrm{F}(\cdot, \vartheta_0)$ zentriert mit echt positiver Varianz ist, folgt $0 < \mathrm{F}(0, \vartheta_0) < 1$. Unter Beachtung von (3.26) konvergiert deshalb die rechten Seite der obigen Ungleichungskette gegen Null. Der Beweis für

$$\mathrm{P}\left(\mathrm{P}\left(\bigcap_{i=1}^{n}\{e_i^* \geq 0\}\Big|\mathcal{A}_n\right) \geq \varepsilon\right) = o(1)$$

verläuft analog, was den Beweis des Lemmas abschließt. □

Damit ist die Menge $\{0 \notin (\min_{1\leq i\leq n} e_i^*, \max_{1\leq i\leq n} e_i^*)\}$ für asymptotische Betrachtungen irrelevant, und auf dem Gegenereignis definieren wir die modifizierte empirische Bootstrap-Verteilungsfunktion

$$\mathrm{F}_{n,z}^*(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+t_n^* e_i^*}1_{\{e_i^* \leq x\}} \qquad \text{für } x \in \mathrm{I\!R}.$$

Dabei ist t_n^* wieder die im Intervall

$$\left(\left(\frac{1}{n}-1\right)\frac{1}{\max_{1\leq i\leq n} e_i^*}, \left(\frac{1}{n}-1\right)\frac{1}{\min_{1\leq i\leq n} e_i^*}\right)$$

eindeutig bestimmte Lösung der Gleichung

$$\sum_{i=1}^{n}\frac{1}{1+t_n^* e_i^*}e_i^* = 0.$$

Als nächstes beweisen wir eine stochastische Entwicklung von t_n^* und $\mathrm{F}_{n,z}^*$ analog zu den Entwicklungen (3.20) und (3.18). Dazu dient das nächste Lemma.

Lemma 4.3

Gelten die Bedingungen (2.5), (4.2) und (3.15), so folgt

$$t_n^* = O_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch,} \tag{4.11}$$

$$t_n^* = \frac{1}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n} e_i^* + o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch} \tag{4.12}$$

und für alle $x \in \mathrm{I\!R}$ gilt

$$\mathrm{F}_{n,z}^*(x) - \mathrm{F}(x, \hat{\vartheta}_{n,res}) = \frac{1}{n}\sum_{i=1}^{n}\left(1_{\{e_i^* \leq x\}} - \mathrm{F}(x, \hat{\vartheta}_{n,res}) - \frac{\mathrm{U}(x)}{\sigma^2}e_i^*\right) + R_n(x) \tag{4.13}$$

mit

$$\sup_{x\in\mathrm{I\!R}} |R_n(x)| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch.}$$

Beweis Zu (4.11):

Wie beim Beweis von (3.19) zeigt man

$$\frac{|t_n^*|}{1+|t_n^*|\max_{1\le i\le n}|e_i^*|}\cdot\frac{1}{n}\sum_{i=1}^n e_i^{*2}\le\left|\frac{1}{n}\sum_{i=1}^n e_i^*\right|$$

und Auflösen nach t_n^* beweist mit (4.9), (4.8) und (4.10) die gewünschte stochastische Ordnung.

Zu (4.12):

Es ist

$$0=\frac{1}{n}\sum_{i=1}^n e_i^*-t_n^*\frac{1}{n}\sum_{i=1}^n e_i^{*2}+t_n^{*2}\frac{1}{n}\sum_{i=1}^n\frac{e_i^{*3}}{1+t_n^*e_i^*}$$

analog zum Beweis von (3.20). Auf Grund von (4.10) und (4.11) ist

$$t_n^*\left(\frac{1}{n}\sum_{i=1}^n e_i^{*2}-\sigma^2\right)=o_{\mathrm{P}^*}(n^{-1/2})\qquad\text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)\ \text{P-stochastisch.}$$

Den letzten Summanden der obigen Gleichung können wir durch

$$\left|t_n^{*2}\frac{1}{n}\sum_{i=1}^n\frac{e_i^{*3}}{1+t_n^*e_i^*}\right|\le t_n^{*2}\cdot\max_{1\le i\le n}\left|\frac{1}{1+t_n^*e_i^*}\right|\cdot\max_{1\le i\le n}|e_i^*|\cdot\frac{1}{n}\sum_{i=1}^n e_i^{*2}$$

abschätzen, und (4.11), (4.9) und (4.10) zeigen, daß der letzte Summand ebenfalls die Ordnung $o_{\mathrm{P}^*}(n^{-1/2})$ unter $\mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)$ P-stochastisch besitzt. Schließlich ergibt Auflösen nach t_n^* die Entwicklung (4.12).

Zu (4.13):

Wir verwenden wieder für jedes $x\in\mathbb{R}$ die Gleichheit $\frac{1}{1+y}=1-y+\frac{y^2}{1+y}$ für $y\neq-1$ und erhalten so

$$\begin{aligned}\mathrm{F}^*_{n,z}(x)-\mathrm{F}(x,\hat\vartheta_{n,res}) &= \frac{1}{n}\sum_{i=1}^n 1_{\{e_i^*\le x\}}-t_n^*\frac{1}{n}\sum_{i=1}^n e_i^*1_{\{e_i^*\le x\}}\\ &\quad+t_n^{*2}\frac{1}{n}\sum_{i=1}^n\frac{e_i^{*2}}{1+t_n^*e_i^*}1_{\{e_i^*\le x\}}-\mathrm{F}(x,\hat\vartheta_{n,res})\\ &= \frac{1}{n}\sum_{i=1}^n\left(1_{\{e_i^*\le x\}}-\mathrm{F}(x,\hat\vartheta_{n,res})-\frac{\mathrm{U}(x)}{\sigma^2}e_i^*\right)\\ &\quad-t_n^*\left(\frac{1}{n}\sum_{i=1}^n e_i^*1_{\{e_i^*\le x\}}-\mathrm{U}(x)\right)\end{aligned}$$

$$
\begin{aligned}
& -\mathrm{U}(x)\left(t_n^* - \frac{1}{n}\frac{1}{\sigma^2}\sum_{i=1}^n e_i^*\right) + t_n^{*\,2}\frac{1}{n}\sum_{i=1}^n \frac{e_i^{*2}}{1+t_n^* e_i^*} 1_{\{e_i^* \le x\}} \\
= \; & \frac{1}{n}\sum_{i=1}^n \left(1_{\{e_i^* \le x\}} - \mathrm{F}(x, \hat\vartheta_{n,res}) - \frac{\mathrm{U}(x)}{\sigma^2} e_i^*\right) \\
& -R_{1n}(x) - R_{2n}(x) + R_{3n}(x).
\end{aligned}
$$

Es reicht, für die drei Restterme die behauptete Ordnung gleichmäßig in $x \in \mathrm{I\!R}$ zu zeigen.

Dies sehen wir sofort für R_{2n}, denn die Funktion U ist beschränkt, und (4.12) impliziert somit

$$
\sup_{x \in \mathrm{I\!R}} |R_{2n}(x)| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)\ \text{ P-stochastisch.}
$$

Andererseits haben wir für R_{3n} die folgende Abschätzung

$$
\sup_{x \in \mathrm{I\!R}} |R_{3n}(x)| \le t_n^{*\,2} \cdot \max_{1 \le i \le n} \frac{1}{1+t_n^* e_i^*} \cdot \frac{1}{n}\sum_{i=1}^n e_i^{*2}.
$$

Dabei ist die rechte Seite unabhängig von x und hat wegen (4.11), (4.9) und (4.10) sogar die Ordnung $O_{\mathrm{P}^*}(n^{-1})$ unter $\mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)$ P-stochastisch.

Schließlich bleibt R_{1n} zu untersuchen. Dazu setzen wir

$$
\mathrm{U}_n^{*\,+}(x) = \frac{1}{n}\sum_{i=1}^n e_i^{*+} 1_{\{e_i^* \le x\}} \qquad \text{für } x \in \mathrm{I\!R}
$$

und definieren $\mathrm{U}_n^{*\,-}$ analog. Dann gilt

$$
\sup_{x \in \mathrm{I\!R}} \left| \frac{1}{n}\sum_{i=1}^n e_i^* 1_{\{e_i^* \le x\}} - \mathrm{U}(x) \right| \le \sup_{x \in \mathrm{I\!R}} \left|\mathrm{U}_n^{*\,+}(x) - \mathrm{U}^+\right| + \sup_{x \in \mathrm{I\!R}} \left|\mathrm{U}_n^{*\,-}(x) - \mathrm{U}^-\right|
$$

mit U^+ und U^- gemäß (3.22), und wegen (4.11) genügt es

$$
\sup_{x \in \mathrm{I\!R}} \left|\mathrm{U}_n^{*\,\pm}(x) - \mathrm{U}^\pm\right| = o_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)\ \text{ P-stochastisch} \tag{4.14}
$$

zu zeigen.

Zunächst stellen wir für jedes $x \in \mathrm{I\!R}$ und alle $\varepsilon' > 0$ fest, daß

$$
\begin{aligned}
\mathrm{P}\left(\left|\mathrm{U}_n^{*\,+}(x) - \mathrm{U}^+(x)\right| \ge \varepsilon' |\mathcal{A}_n\right) \;\le\; & \mathrm{P}\left(\left|\mathrm{U}_n^{*\,+}(x) - \mathrm{E}\left(e_1^{*+} 1_{\{e_1^* \le x\}} |\mathcal{A}_n\right)\right| \ge \frac{\varepsilon'}{2} \Big| \mathcal{A}_n\right) \\
& + 1_{\left\{\left|\mathrm{E}\left(e_1^{*+} 1_{\{e_1^* \le x\}}|\mathcal{A}_n\right) - \mathrm{U}^+(x)\right| \ge \frac{\varepsilon'}{2}\right\}}
\end{aligned}
$$

gilt, und der zweite Summand auf der rechten Seite der Ungleichung konvergiert wegen (4.1) und (4.5) bereits stochastisch gegen Null. Für den ersten gilt aber auf Grund der Tschbyscheff-Ungleichung

$$\mathrm{P}\left(\left|\mathrm{U}_n^{*+}(x) - \mathrm{E}\left(e_1^{*+}1_{\{e_1^*\leq x\}}|\mathcal{A}_n\right)\right| \geq \frac{\varepsilon'^2}{2}\Big|\mathcal{A}_n\right) \leq \frac{4}{\varepsilon'^2 n}\mathrm{E}\left(e_1^{*2}|\mathcal{A}_n\right),$$

und der Term auf der rechten Seite konvergiert wegen (4.10) stochastisch gegen Null. Für jedes feste $x \in \mathbb{R}$ gilt also

$$\left|\mathrm{U}_n^{*+}(x) - \mathrm{U}^+(x)\right| = o_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\left(\,\cdot\,|\mathcal{A}_n\right)\ \text{P-stochastisch,}$$

und genauso zeigt man

$$\left|\mathrm{U}_n^{*-}(x) - \mathrm{U}^-(x)\right| = o_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\left(\,\cdot\,|\mathcal{A}_n\right)\ \text{P-stochastisch.}$$

Um nun die Aussage (4.14) zu beweisen, benutzen wir ein Monotonieargument. Setzen wir U^+ durch

$$\mathrm{U}^+(-\infty) = 0 \qquad \text{und} \qquad \mathrm{U}^+(\infty) = \mathrm{E}(e_1^+)$$

fort, so ist U^+ auf $[-\infty, \infty]$ gleichmäßig stetig und beschränkt. Zu jedem $\varepsilon > 0$ existieren deshalb Punkte $-\infty = x_0 < x_1 < \ldots < x_m < x_{m+1} = \infty$, so daß für alle $k \in \{0, \ldots, m\}$ die Abschätzung

$$\mathrm{U}^+(x_{k+1}) - \mathrm{U}^+(x_k) < \varepsilon$$

gilt. Für jedes $x \in \mathbb{R}$ existiert nun genau ein $k \in \{0, \ldots, m\}$ mit $x_k \leq x < x_{k+1}$, und es gilt

$$\mathrm{U}_n^{*+}(x) - \mathrm{U}^+(x) \leq \mathrm{U}_n^{*+}(x_{k+1})) - \mathrm{U}^+(x_{k+1}) + \varepsilon$$

auf Grund der Monotonie von U_n^{*+} und U^+. Ebenso erhalten wir

$$\mathrm{U}^+(x) - \mathrm{U}_n^{*+}(x) \leq \mathrm{U}^+(x_k) - \mathrm{U}_n^{*+}(x_k) + \varepsilon$$

und damit folgt

$$\sup_{x\in\mathbb{R}}\left|\mathrm{U}_n^{*+}(x) - \mathrm{U}^+(x)\right| \leq \varepsilon + \max_{0\leq k\leq m+1}\left|\mathrm{U}_n^{*+}(x_k) - \mathrm{U}^+(x_k)\right|.$$

Wie oben gesehen, hat das endliche Maximum auf der rechten Seite der Ungleichung die behauptete Ordnung, und da $\varepsilon > 0$ beliebig klein ist, folgt

$$\sup_{x\in\mathbb{R}}\left|\mathrm{U}_n^{*+}(x) - \mathrm{U}^+\right| = o_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\left(\,\cdot\,|\mathcal{A}_n\right)\ \text{P-stochastisch.}$$

Die entsprechende Aussage für die Negativteile zeigt man analog, woraus (4.14) und somit auch

$$\sup_{x\in\mathbb{R}}|R_{1n}(x)| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\left(\,\cdot\,|\mathcal{A}_n\right)\ \text{P-stochastisch}$$

folgt.

Damit sind alle Teile des Lemmas bewiesen. □

4.2.3 Der Bootstrap-Schätzer für den Verteilungsparameter

In diesem Abschnitt untersuchen wir die Bootstrap-Version zur Verteilungsfunktion $\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})$. Dazu benötigen wir zunächst einen Bootstrap-Schätzer für den Verteilungsparameter. Es ist klar, daß dieser eine (3.15) entsprechende Entwicklung haben sollte. Dabei müssen die Voraussetzungen an die Einflußfunktion aus (2.6) im lokal gleichmäßigen Sinne verschärft werden. Dies führt zu folgenden Bedingungen:

Die Einflußfunktion L ist für alle $\vartheta \in V$ mit V aus (4.2) definiert, und für $1 \le k, l \le d$ gelten

$$\int_{\mathbb{R}} (\mathrm{L}(x,\vartheta))_k \, \mathrm{F}(dx,\vartheta) = 0 \qquad \text{für } \vartheta \in V$$

und

$$\int_{\mathbb{R}} \left(\mathrm{L}(x,\vartheta)\mathrm{L}(x,\vartheta)^t\right)_{kl} \mathrm{F}(dx,\vartheta) \longrightarrow \mathrm{E}\left(\mathrm{L}(e_1,\vartheta_0)\mathrm{L}(e_1,\vartheta_0)^t\right)_{kl} \tag{4.15}$$

für $\vartheta \to \vartheta_0$.

Es ist $(\hat{\vartheta}^*_n)_{n\in\mathbb{N}}$ eine Schätzerfolge mit

$$\hat{\vartheta}^*_n - \hat{\vartheta}_{n,res} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{L}(e^*_i, \hat{\vartheta}_{n,res}) + R_n,$$

wobei

$$\|R_n\| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch} \tag{4.16}$$

gilt.

Gilt

$$\hat{\vartheta}_{n,res} = \mathrm{p}_n(\hat{e}_1, \ldots, \hat{e}_n)$$

mit einer bekannten Folge von Funktionen $\mathrm{p}_n : \mathbb{R}^n \longrightarrow \mathbb{R}^d$, so kann man

$$\hat{\vartheta}^*_n = \mathrm{p}_n(e^*_1, \ldots, e^*_n)$$

setzen. Für den so definierten Bootstrap-Schätzer wäre dann die Entwicklung (4.16) im Einzelfall nachzuweisen.

Andererseits kann die Bedingung (4.16) trivial erfüllt werden, wenn die Einflußfunktion L wie im Falle der schon betrachteten Maximum-Likelihood-Schätzung bekannt ist. Wir können in diesem Falle den Schätzer $\hat{\vartheta}^*_n$ durch

$$\hat{\vartheta}^*_n = \hat{\vartheta}_{n,res} + \frac{1}{n}\sum_{i=1}^{n} \mathrm{L}(e^*_i, \hat{\vartheta}_{n,res})$$

definieren und erhalten somit die Entwicklung (4.16) sogar mit trivialem Restterm.

Den oben eingeführten Schätzer $\hat{\vartheta}_n^*$ verwenden wir nun zur Definition von $\mathrm{F}(\cdot, \hat{\vartheta}_n^*)$, einer Bootstrap-Version von $\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})$. Dazu müssen wir sicherstellen, daß $\hat{\vartheta}_n^* \in \Theta$ gilt.

Lemma 4.4

Unter den Bedingungen (3.15), (4.15) und (4.16) gilt

$$\mathrm{P}\left(\hat{\vartheta}_n^* \notin V \middle| \mathcal{A}_n\right) \longrightarrow 0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Insbesondere ist $\{\hat{\vartheta}_n^* \notin \Theta\}$ für Verteilungskonvergenzaussagen asymptotisch vernachlässigbar.

Beweis Es existiert ein $\varepsilon_0' > 0$, so daß $\{\vartheta \in \Theta : \|\vartheta - \vartheta_0\| < \varepsilon_0'\} \subseteq V$ gilt. Daraus folgern wir

$$\mathrm{P}\left(\hat{\vartheta}_n^* \notin V \middle| \mathcal{A}_n\right) \leq \mathrm{P}\left(\left\|\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res}\right\| \geq \frac{\varepsilon_0'}{2} \middle| \mathcal{A}_n\right) + 1_{\left\{\left\|\hat{\vartheta}_{n,res} - \vartheta_0\right\| \geq \frac{\varepsilon_0'}{2}\right\}},$$

wobei der zweite Summand auf der rechten Seite der Ungleichung wegen (3.15) eine stochastische Nullfolge bildet. Für den ersten Summanden gilt wegen (4.16) und der stochastischen Ordnung des dort betrachteten Restterms

$$\mathrm{P}\left(\left\|\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res}\right\| \geq \frac{\varepsilon_0'}{2} \middle| \mathcal{A}_n\right) \leq \mathrm{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res})\right\| \geq \frac{\varepsilon_0'}{4} \middle| \mathcal{A}_n\right) + o_{\mathrm{P}}(1).$$

Es genügt also, die stochastische Konvergenz der bedingten Wahrscheinlichkeit auf der rechten Seite gegen Null zu zeigen. Wir werden später noch die stärkere stochastische Rate

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res}) = O_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch,} \tag{4.17}$$

benötigen, weshalb wir schon hier diese Aussage beweisen. Wegen (4.15) gilt für jedes $C > 0$ die Ungleichung

$$\begin{aligned} &\mathrm{P}\left(\left\{\frac{1}{\sqrt{n}}\left\|\sum_{i=1}^n \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res})\right\| \geq C\right\} \cap \left\{\left\|\hat{\vartheta}_{n,res} - \vartheta_0\right\| < \varepsilon_0'\right\} \middle| \mathcal{A}_n\right) \\ \leq\ & \frac{1}{C^2 n}\mathrm{E}\left(\left\|\sum_{i=1}^n \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res})\right\|^2 \middle| \mathcal{A}_n\right) \cdot 1_{\left\{\left\|\hat{\vartheta}_{n,res} - \vartheta_0\right\| < \varepsilon_0'\right\}} \\ =\ & \frac{1}{C^2}\sum_{k=1}^d \mathrm{E}\left(\mathrm{L}_k^2(e_1^*, \hat{\vartheta}_{n,res}) | \mathcal{A}_n\right) \cdot 1_{\left\{\left\|\hat{\vartheta}_{n,res} - \vartheta_0\right\| < \varepsilon_0'\right\}}. \end{aligned}$$

Wählen wir nun zu gegebenem $\varepsilon > 0$ die Schranke $C > 0$ so, daß

$$\max_{1\leq k\leq d} \mathrm{E}\left(\mathrm{L}_k^2(e_1, \vartheta_0)\right) < \frac{C^2\varepsilon}{4d}$$

gilt, so erhalten wir

$$\begin{aligned}
&\mathrm{P}\left(\mathrm{P}\left(\frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n} \mathrm{L}_(e_i^*, \hat{\vartheta}_{n,res})\right\| \geq C\middle|\mathcal{A}_n\right) \geq \varepsilon\right) \\
&\leq \ \mathrm{P}\left(\sum_{k=1}^{d} \mathrm{E}\left(\mathrm{L}_k^2(e_1^*, \hat{\vartheta}_{n,res})\middle|\mathcal{A}_n\right) \cdot 1_{\left\{\left\|\hat{\vartheta}_{n,res}-\vartheta_0\right\|<\varepsilon_0'\right\}} \geq \frac{\varepsilon C^2}{2}\right) \\
&\qquad +\mathrm{P}\left(\left\|\hat{\vartheta}_{n,res} - \vartheta_0\right\| \geq \varepsilon_0'\right) \\
&\leq \ \sum_{k=1}^{d} \mathrm{P}\left(\mathrm{E}\left(\mathrm{L}_k^2(e_1^*, \hat{\vartheta}_{n,res})\middle|\mathcal{A}_n\right) \cdot 1_{\left\{\left\|\hat{\vartheta}_{n,res}-\vartheta_0\right\|<\varepsilon_0'\right\}} \geq \frac{\varepsilon C^2}{2d}\right) + o(1) \\
&\leq \ \sum_{k=1}^{d} \mathrm{P}\left(\left|\mathrm{E}\left(\mathrm{L}_k^2(e_1^*, \hat{\vartheta}_{n,res})\middle|\mathcal{A}_n\right) \cdot 1_{\left\{\left\|\hat{\vartheta}_{n,res}-\vartheta_0\right\|<\varepsilon_0'\right\}} - \mathrm{E}\left(\mathrm{L}_k^2(e_1, \vartheta_0)\right)\right| \geq \frac{\varepsilon C^2}{4d}\right) \\
&\qquad +o(1),
\end{aligned}$$

und die rechte Seite der Ungleichung konvergiert wegen (4.1) und (4.15) gegen Null, was (4.17) beweist. Damit ist das Lemma gezeigt. □

Mit $\mathrm{F}(\cdot, \hat{\vartheta}_n^*)$ haben wir nun einen Bootstrap-Schätzer für $\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})$ gefunden. Es ist klar, daß wir wieder gleichmäßig in $x \in \mathbb{R}$ für die Differenz $\mathrm{F}(x, \hat{\vartheta}_n^*) - \mathrm{F}(x, \hat{\vartheta}_{n,res})$ eine stochastische Entwicklung herleiten müssen. Da im Bootstrap-Kontext der „wahre“ Verteilungsparameter $\hat{\vartheta}_{n,res}$ zufällig ist, ist (2.5) für die angestrebte stochastische Entwicklung nicht mehr hinreichend. Deshalb ersetzen wir diese Voraussetzung durch die folgende lokal gleichmäßige Version.

> Für alle $\vartheta \in V$ ist die Funktion $\Delta(\cdot, \vartheta) : \mathbb{R} \longrightarrow \mathbb{R}^d$ definiert, $\Delta(\cdot, \vartheta_0)$ ist stetig im ersten Argument, und es gilt
>
> $$\sup_{\substack{x\in\mathbb{R}\\ \vartheta\in V}} \left|\mathrm{F}(x, \vartheta + h) - \mathrm{F}(x, \vartheta)) - \Delta(x, \vartheta)^t h\right| = o(\|h\|) \qquad \text{für } \|h\| \to 0. \tag{4.18}$$

Außerdem benötigen wir für das folgende Lemma die Glattheitsvoraussetzung

$$\sup_{x\in\mathbb{R}} \|\Delta(x, \vartheta) - \Delta(x, \vartheta_0)\| \longrightarrow 0 \qquad \text{für } \vartheta \to \vartheta_0. \tag{4.19}$$

Lemma 4.5

Gelten die Voraussetzungen (3.15), (4.15), (4.16), (4.18) und (4.19), so gilt für $x \in \mathbb{R}$ die Entwicklung

$$\mathrm{F}(x, \hat{\vartheta}_n^*) - \mathrm{F}(x, \hat{\vartheta}_{n,res}) = \Delta(x, \vartheta_0)^t \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res}) + R_n(x) \tag{4.20}$$

mit

$$\sup_{x \in \mathbb{R}} |R_n(x)| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)\ \text{ P-stochastisch.}$$

Beweis Für jedes $x \in \mathbb{R}$ ist

$$\begin{aligned}
\mathrm{F}(x, \hat{\vartheta}_n^*) - \mathrm{F}(x, \hat{\vartheta}_{n,res}) &= \Delta(x, \vartheta_0)^t \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res}) \\
&\quad + \Big(\mathrm{F}(x, \hat{\vartheta}_n^*) - \mathrm{F}(x, \hat{\vartheta}_{n,res}) - \Delta(x, \hat{\vartheta}_{n,res})^t \left(\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res}\right)\Big) \\
&\quad + \Delta(x, \hat{\vartheta}_{n,res})^t \left(\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res} - \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res})\right) \\
&\quad + \Big(\Delta(x, \hat{\vartheta}_{n,res}) - \Delta(x, \vartheta_0)\Big)^t \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res}) \\
&= \Delta(x, \vartheta_0)^t \frac{1}{n} \sum_{i=1}^{n} \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res}) \\
&\quad + R_{1n}(x) + R_{2n}(x) + R_{3n}(x),
\end{aligned}$$

und es reicht aus, für $s \in \{1, 2, 3\}$ zu zeigen, daß

$$\sup_{x \in \mathbb{R}} |R_{sn}(x)| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)\ \text{ P-stochastisch}$$

gilt. Zur Diskussion von R_{1n} seien $\varepsilon, \varepsilon' > 0$ vorgegeben. Wegen (4.16) und (4.17) existiert ein $C > 0$ so groß, daß

$$\limsup_{n \to \infty} \mathrm{P}\left(\mathrm{P}\left(\sqrt{n}\left\|\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res}\right\| > C\middle|\mathcal{A}_n\right) \geq \frac{\varepsilon}{3}\right) \leq \varepsilon \tag{4.21}$$

gilt. Wegen der Voraussetzung (4.18) gibt es weiter zu $\varepsilon', C > 0$ eine Zahl $h_0 > 0$, so daß

$$\sup_{\substack{x \in \mathbb{R} \\ \vartheta \in V}} \left|\mathrm{F}(x, \vartheta + h) - \mathrm{F}(x, \vartheta)) - \Delta(x, \vartheta)^t h\right| \leq \|h\| \frac{\varepsilon'}{2C} \tag{4.22}$$

für alle $h \in \mathbb{R}^d$ mit $\|h\| \le h_0$ gilt. Wir schätzen nun

$$\begin{aligned}
&\mathrm{P}\left(\sqrt{n}\sup_{x\in\mathbb{R}}|R_{1n}(x)| \ge \varepsilon' \Big| \mathcal{A}_n\right)\\
\le\ & \mathrm{P}\left(\left\{\sqrt{n}\sup_{x\in\mathbb{R}}|R_{1n}(x)| \ge \varepsilon'\right\} \cap \left\{\sqrt{n}\left\|\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res}\right\| \le C\right\} \cap \left\{\hat{\vartheta}_{n,res} \in V\right\} \Big| \mathcal{A}_n\right)\\
& + \mathrm{P}\left(\sqrt{n}\left\|\hat{\vartheta}_n^* - \hat{\vartheta}_{n,res}\right\| > C \Big| \mathcal{A}_n\right) + 1_{\{\hat{\vartheta}_{n,res}\notin V\}}\\
=\ & P_{1n} + P_{2n} + P_{3n}
\end{aligned}$$

ab. Für alle großen n ist $\frac{C}{\sqrt{n}} < h_0$ und deshalb mit (4.22) die Menge in P_{1n} leer. Weiter ist wegen (4.1) auch

$$\limsup_{n\to\infty} \mathrm{P}\left(P_{3n} \ge \frac{\varepsilon}{3}\right) \le \limsup_{n\to\infty} \mathrm{P}\left(\hat{\vartheta}_{n,res} \notin V\right) = 0.$$

Deshalb können wir

$$\limsup_{n\to\infty} \mathrm{P}\left(\mathrm{P}\left(\sqrt{n}\sup_{x\in\mathbb{R}}|R_{1n}(x)| \ge \varepsilon' \Big| \mathcal{A}_n\right) \ge \varepsilon\right) \le \limsup_{n\to\infty} \mathrm{P}\left(P_{2n} \ge \frac{\varepsilon}{3}\right) \le \varepsilon$$

folgern, wobei wir (4.21) verwenden. Da $\varepsilon > 0$ beliebig ist, schließt dies den Beweis für R_{1n} ab.

Damit wenden wir uns R_{2n} zu. Wegen (4.16) reicht es,

$$\sup_{x\in\mathbb{R}} \|\Delta(x, \hat{\vartheta}_{n,res})\| = O_{\mathrm{P}}(1)$$

zu zeigen. Wegen (4.18) existiert ein $h_0 > 0$, so daß

$$\sup_{\substack{x\in\mathbb{R}\\ \vartheta\in V}} \left|\mathrm{F}(x, \vartheta + h) - \mathrm{F}(x, \vartheta) - \Delta(x, \vartheta)^t h\right| \le \|h\| \qquad \text{für } \|h\| \le h_0$$

ist. Bezeichnen wir nun für $1 \le k \le d$ mit v_k den k-ten Einheitsvektor, so gilt für beliebiges $\vartheta \in V$ die Ungleichung

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |\Delta(x, \vartheta)_k| &= \frac{1}{h_0}\sup_{x\in\mathbb{R}} |-\Delta(x, \vartheta)_k h_0|\\
&\le \frac{1}{h_0}\sup_{x\in\mathbb{R}} |\mathrm{F}(x, \vartheta + h_0 v_k) - \mathrm{F}(x, \vartheta) - \Delta(x, \vartheta)^t h_0 v_k|\\
&\quad + \frac{1}{h_0}\sup_{x\in\mathbb{R}} |\mathrm{F}(x, \vartheta + h_0 v_k) - \mathrm{F}(x, \vartheta)|\\
&\le 1 + \frac{1}{h_0}
\end{aligned}$$

und somit

$$\sup_{\substack{x\in\mathbb{R}\\ \vartheta\in V}} \|\Delta(x, \vartheta)\| \le \sqrt{d}\left(1 + \frac{1}{h_0}\right) < \infty,$$

was für alle großen C die Abschätzung

$$\limsup_{n\to\infty} \mathrm{P}\left(\left\|\Delta(x,\hat{\vartheta}_{n,res})\right\| \geq C\right) \leq \limsup_{n\to\infty} \mathrm{P}\left(\hat{\vartheta}_{n,res} \notin V\right) = 0$$

gemäß (4.1) ergibt. Damit erhalten wir die behauptete stochastische Ordnung für R_{2n}.

Zuletzt diskutieren wir das asymptotische Verhalten von R_{3n}. Hierzu stellen wir zunächst fest, daß es wegen (4.17) genügt,

$$\sup_{x\in\mathbb{R}} \left\|\Delta(x,\hat{\vartheta}_{n,res}) - \Delta(x,\vartheta_0)\right\| = o_{\mathrm{P}}(1)$$

zu zeigen. Auf Grund von (4.19) existiert zu gegebenem $\varepsilon > 0$ ein $\delta > 0$ mit

$$\sup_{x\in\mathbb{R}} \|\Delta(x,\vartheta) - \Delta(x,\vartheta_0)\| \leq \frac{\varepsilon}{2} \qquad \text{für } \|\vartheta - \vartheta_0\| \leq \delta,$$

und somit ist

$$\limsup_{n\to\infty} \mathrm{P}\left(\left\|\Delta(x,\hat{\vartheta}_{n,res}) - \Delta(x,\vartheta_0)\right\| \geq \varepsilon\right) \leq \limsup_{n\to\infty} \mathrm{P}\left(\left\|\hat{\vartheta}_{n,res} - \vartheta_0\right\| > \delta\right) = 0$$

gemäß (4.1). Dies zeigt die gewünschte Ordnung für R_{3n}, was den Beweis des Lemmas abschließt. □

4.2.4 Der modifizierte empirische Bootstrap-Prozeß mit geschätztem Parameter

Wir sind nun in der Lage, das asymptotische Verteilungsverhalten des Prozesses $(n^{1/2}(\mathrm{F}^*_{n,z} - \mathrm{F}(x,\hat{\vartheta}^*_n)))_{n\in\mathbb{N}}$ zu studieren. Dazu ist wiederum eine stochastische Entwicklung grundlegend.

Lemma 4.6

Unter den Voraussetzungen (3.15), (4.2), (4.15), (4.16), (4.18) und (4.19) gilt

$$\mathrm{F}^*_{n,z}(x) - \mathrm{F}(x,\hat{\vartheta}^*_n) = \frac{1}{n}\sum_{i=1}^{n} Y^*_i(x) + R_n(x) \qquad \text{für } x \in \mathbb{R} \tag{4.23}$$

mit

$$Y^*_i(x) = 1_{\{e^*_i \leq x\}} - \mathrm{F}(x,\hat{\vartheta}_{n,res}) - \frac{\mathrm{U}(x)}{\sigma^2}e^*_i - \Delta(x,\vartheta_0)^t \mathrm{L}(e^*_i,\hat{\vartheta}^*_n)$$

und

$$\sup_{x\in\mathbb{R}} |R_n(x)| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n)\ \text{P-stochastisch.}$$

Beweis Dies folgt direkt aus (4.13) und (4.20). □

Als nächstes werden wir einen funktionalen Grenzwertsatz für den stochastischen Prozeß $(n^{1/2}(\mathrm{F}^*_{n,z} - \mathrm{F}(\cdot, \hat{\vartheta}^*_n)))_{n\in\mathbb{N}}$ beweisen. Dazu wiederholen wir nochmals die bisher benötigten Voraussetzungen.

Es existiert eine meßbare Funktion $\mathrm{L}(\cdot, \vartheta_0) : \mathbb{R} \longrightarrow \mathbb{R}^d$ mit Erwartungswertvektor $\mathrm{E}(\mathrm{L}(e_1, \vartheta_0)) = 0$, existierender Kovarianzmatrix $\Lambda(\vartheta_0)$ und

$$\hat{\vartheta}_{n,res} - \vartheta_0 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{L}(e_i, \vartheta_0) + R_n. \tag{3.15}$$

Dabei gilt $\|R_n\| = o_{\mathrm{P}}(n^{-1/2})$.

Es existiert eine abgeschlossene Umgebung V von ϑ_0, so daß

$$\int_{\mathbb{R}} x\mathrm{F}(dx, \vartheta) = 0 \qquad \text{für } \vartheta \in V$$

und

$$\int_{\mathbb{R}} x^2\mathrm{F}(dx, \vartheta) \longrightarrow \int_{\mathbb{R}} x^2\mathrm{F}(dx, \vartheta_0) = \sigma^2 \qquad \text{für } \vartheta \to \vartheta_0 \tag{4.2}$$

gilt.

Die Einflußfunktion L ist für alle $\vartheta \in V$ mit V aus (4.2) definiert, und für $1 \leq k, l \leq d$ gelten

$$\int_{\mathbb{R}} (\mathrm{L}(x, \vartheta))_k \, \mathrm{F}(dx, \vartheta) = 0 \qquad \text{für } \vartheta \in V$$

und

$$\int_{\mathbb{R}} \left(\mathrm{L}(x, \vartheta)\mathrm{L}(x, \vartheta)^t\right)_{kl} \mathrm{F}(dx, \vartheta) \longrightarrow \mathrm{E}\left(\mathrm{L}(e_1, \vartheta_0)\mathrm{L}(e_1, \vartheta_0)^t\right)_{kl} \tag{4.15}$$

für $\vartheta \to \vartheta_0$.

Es ist $(\hat{\vartheta}^*_n)_{n\in\mathbb{N}}$ eine Schätzerfolge mit

$$\hat{\vartheta}^*_n - \hat{\vartheta}_{n,res} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{L}(e^*_i, \hat{\vartheta}_{n,res}) + R_n,$$

wobei

$$\|R_n\| = o_{\mathrm{P}^*}(n^{-1/2}) \qquad \text{unter } \mathrm{P}\,(\,\cdot\,|\mathcal{A}_n) \text{ P-stochastisch} \tag{4.16}$$

gilt.

Für alle $\vartheta \in V$ ist die Funktion $\Delta(\cdot,\vartheta) : \mathbb{R} \longrightarrow \mathbb{R}^d$ definiert, $\Delta(\cdot,\vartheta_0)$ ist stetig im ersten Argument, und es gilt

$$\sup_{\substack{x\in\mathbb{R}\\ \vartheta\in V}} \left|\mathrm{F}(x,\vartheta+h) - \mathrm{F}(x,\vartheta)) - \Delta(x,\vartheta)^t h\right| = o(\|h\|) \qquad \text{für } \|h\| \to 0 \tag{4.18}$$

und

$$\sup_{x\in\mathbb{R}} \|\Delta(x,\vartheta) - \Delta(x,\vartheta_0)\| \longrightarrow 0 \qquad \text{für } \vartheta \to \vartheta_0. \tag{4.19}$$

Wie wir schon vorher gesehen haben, müssen wir in der Situation des Bootstraps einige Regularitätsvoraussetzungen gegenüber der Situation des nicht gebootstrappten Prozesses verschärfen, beziehungsweise neu einführen.

Für $\vartheta \in V$, $1 \le k \le d$ und beliebiges $x \in \mathbb{R}$ gilt

$$\int_{\mathbb{R}} (\mathrm{L}(y,\vartheta))_k 1_{\{y\le x\}} \mathrm{F}(dy,\vartheta) \longrightarrow \int_{\mathbb{R}} (\mathrm{L}(y,\vartheta_0))_k 1_{\{y\le x\}} \mathrm{F}(dy,\vartheta_0) \tag{4.24}$$

für $\vartheta \to \vartheta_0$.

Für $\vartheta \in V$, $1 \le k \le d$ gilt

$$\int_{\mathbb{R}} y\mathrm{L}_k(y,\vartheta)\mathrm{F}(dy,\vartheta) \longrightarrow \int_{\mathbb{R}} y\mathrm{L}_k(y,\vartheta_0)\mathrm{F}(dy,\vartheta_0) \qquad \text{für } \vartheta \to \vartheta_0. \tag{4.25}$$

Für $\vartheta \in V$, $1 \le k \le d$ und alle $C > 0$ gilt

$$\int_{\mathbb{R}} \mathrm{L}_k(y,\vartheta)^2 1_{\{\mathrm{L}_k(y,\vartheta)^2\ge C\}} \mathrm{F}(dy,\vartheta) \longrightarrow \int_{\mathbb{R}} \mathrm{L}_k(y,\vartheta_0)^2 1_{\{\mathrm{L}_k(y,\vartheta_0)^2\ge C\}} \mathrm{F}(dy,\vartheta_0) \tag{4.26}$$

für $\vartheta \to \vartheta_0$.

Wir beweisen nun den folgenden zu (3.27) analogen Grenzwertsatz für den modifizierten empirischen Bootstrap-Prozeß mit geschätztem Parameter.

Satz 4.7

Gelten die obigen Bedingungen, also (3.15), (4.2), (4.15), (4.16), (4.18), (4.19), (4.24), (4.25) und (4.26), so gilt

$$\sqrt{n}\left(\mathrm{F}^*_{n,z} - \mathrm{F}(\cdot,\hat{\vartheta}^*_n)\right) \xrightarrow{\mathcal{L}} \mathrm{W} \qquad \text{in } \mathrm{D}[-\infty,\infty] \qquad \text{für } n \to \infty \tag{4.27}$$

unter $\mathrm{P}(\,\cdot\,|\mathcal{A}_n)$ P-stochastisch. Dabei ist W der Gaußprozeß aus (2.19) und (3.27).

Beweis Zunächst stellen wir fest, daß es gemäß der Entwicklung (4.23) ausreicht,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i^* \xrightarrow{\mathcal{L}} \mathrm{W} \qquad \text{in } \mathrm{D}[-\infty,\infty] \qquad \text{für } n\to\infty \tag{4.28}$$

unter $\mathrm{P}(\,\cdot\,|\mathcal{A}_n)$ P-stochastisch zu zeigen. Wie in Kapitel 2 beweisen wir dazu die Konvergenz der endlich dimensionalen Randverteilungen und die C-Straffheit des Prozesses.

Zuerst beweisen wir die Konvergenz der endlichdimensionalen Randverteilungen. Sei also eine endliche Menge von Punkten $\{t_1,\ldots,t_m\}$ aus $[-\infty,\infty]$ vorgegeben und

$$\underline{Y}_i^*(\underline{t}_m) = (Y_i^*(t_1),\ldots,Y_i^*(t_m))^t$$

gesetzt. Die $\underline{Y}_i^*(\underline{t}_m)$ bilden ein zeilenweise unabhängiges Dreiecksschema bedingt zentrierter und quadratisch integrierbarer Zufallsvektoren. Deshalb können wir eine Bootstrapversion des zentralen Grenzwertsatzes für zeilenweise unabhängige Dreiecksschemata anwenden, sofern wir die Normierungsbedingung und die Lindebergbedingung zeigen können.

Für beliebige $k,l\in\{1,\ldots,m\}$ gilt

$$\begin{aligned}
&\sum_{i=1}^{n}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\underline{Y}_i^*(\underline{t}_m)\middle|\mathcal{A}_n\right)_{kl} = \mathrm{E}\left(Y_1^*(t_k)Y_1^*(t_l)|\mathcal{A}_n\right)\\
=\;& \mathrm{F}\left(\min\{t_k,t_l\},\hat{\vartheta}_{n,res}\right)-\mathrm{F}\left(t_k,\hat{\vartheta}_{n,res}\right)\mathrm{F}\left(t_l,\hat{\vartheta}_{n,res}\right)\\
-\;& \frac{\mathrm{U}(t_k)}{\sigma^2}\mathrm{E}\left(e_1^*1_{\{e_1^*\le t_l\}}|\mathcal{A}_n\right)-\frac{\mathrm{U}(t_l)}{\sigma^2}\mathrm{E}\left(e_1^*1_{\{e_1^*\le t_k\}}|\mathcal{A}_n\right)+\frac{\mathrm{U}(t_k)\mathrm{U}(t_l)}{\sigma^4}\mathrm{E}\left(e_1^{*2}|\mathcal{A}_n\right)\\
-\;& \Delta(t_k,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})1_{\{e_1^*\le t_l\}}\middle|\mathcal{A}_n\right)-\Delta(t_l,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})1_{\{e_1^*\le t_k\}}\middle|\mathcal{A}_n\right)\\
+\;& \frac{\mathrm{U}(t_k)}{\sigma^2}\Delta(t_l,\vartheta_0)^t\mathrm{E}\left(e_1^*\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})\middle|\mathcal{A}_n\right)+\frac{\mathrm{U}(t_l)}{\sigma^2}\Delta(t_k,\vartheta_0)^t\mathrm{E}\left(e_1^*\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})\middle|\mathcal{A}_n\right)\\
+\;& \Delta(t_k,\vartheta_0)^t\mathrm{E}\left(\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})^t\middle|\mathcal{A}_n\right)\Delta(t_l,\vartheta_0),
\end{aligned}$$

wobei wir $\mathrm{E}(e_1^*|\mathcal{A}_n)=0$ und $\mathrm{E}(\mathrm{L}(e_1^*,\hat{\vartheta}_{n,res})|\mathcal{A}_n)=0$ verwenden, was aus (4.2) und (4.15) gefolgert werden kann, wenn wir ohne Einschränkung $\hat{\vartheta}_{n,res}\in V$ annehmen.

Auf Grund von (4.1), (4.3), (4.5), (4.2), (4.24), (4.25) und (4.15) gelten

$$\mathrm{F}(t_k,\hat{\vartheta}_{n,res}) \longrightarrow \mathrm{F}(t_k,\vartheta_0) \qquad \text{für } n\to\infty \qquad \text{P-stochastisch,}$$

$$\mathrm{E}\left(e_1^*1_{\{e_1^*\le t_k\}}|\mathcal{A}_n\right) \longrightarrow \mathrm{U}(t_k) \qquad \text{für } n\to\infty \qquad \text{P-stochastisch,}$$

$$\mathrm{E}\left(e_1^{*2}|\mathcal{A}_n\right) \longrightarrow \sigma^2 \qquad \text{für } n\to\infty \qquad \text{P-stochastisch}$$

und so fort, so daß wir

$$\sum_{i=1}^{n} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\underline{Y}_i^*(\underline{t}_m)\Big|\mathcal{A}_n\right)_{kl} \longrightarrow \Gamma_{kl} \qquad \text{für } n \to \infty \qquad \text{P-stochastisch}$$

folgern können. Dabei ist die Matrix Γ durch

$$\Gamma_{kl} = \mathrm{Cov}(\mathrm{W}(t_k), \mathrm{W}(t_l)) \qquad \text{für } 1 \leq k, l \leq m$$

definiert. Die bedingte Varianz der endlichdimensionalen Projektionen des Prozesses $(n^{-1/n}\sum_{i=1}^{n} Y_i^*)_{n\in\mathbb{N}}$ konvergiert also stochastisch gegen die gewünschte Grenzvarianz, was bedeutet, daß die Normierungsbedingung erfüllt ist.

Zum Nachweis der Lindebergbedingung muß für beliebiges $\varepsilon > 0$ die stochastische Ordnung

$$\sum_{i=1}^{n} \mathrm{E}\left(\left\|\frac{1}{\sqrt{n}}\underline{Y}_i^*(\underline{t}_m)\right\|^2 1_{\left\{\|\underline{Y}_i^*(\underline{t}_m)\|\geq\varepsilon\sqrt{n}\right\}}\Big|\mathcal{A}_n\right) = o_{\mathrm{P}}(1)$$

gezeigt werden. Wegen der zeilenweise identischen Verteiltheit gilt

$$\begin{aligned}
&\sum_{i=1}^{n} \mathrm{E}\left(\left\|\frac{1}{\sqrt{n}}\underline{Y}_i^*(\underline{t}_m)\right\|^2 1_{\left\{\|\underline{Y}_i^*(\underline{t}_m)\|\geq\varepsilon\sqrt{n}\right\}}\Big|\mathcal{A}_n\right) \\
&\qquad = \mathrm{E}\left(\|\underline{Y}_1^*(\underline{t}_m)\|^2\, 1_{\left\{\|\underline{Y}_1^*(\underline{t}_m)\|\geq\varepsilon\sqrt{n}\right\}}\Big|\mathcal{A}_n\right) \\
&\qquad \leq \sum_{k=1}^{m}\sum_{l=1}^{m} \mathrm{E}\left(Y_1^*(t_k)^2 1_{\left\{Y_1^*(t_l)^2\geq\frac{\varepsilon^2 n}{m}\right\}}\Big|\mathcal{A}_n\right),
\end{aligned}$$

und es reicht, für beliebige $k, l \in \{1, \ldots m\}$ lediglich

$$\mathrm{E}\left(Y_1^*(t_k)^2 1_{\left\{Y_1^*(t_l)^2\geq\frac{\varepsilon^2 n}{m}\right\}}\Big|\mathcal{A}_n\right) = o_{\mathrm{P}}(1)$$

zu beweisen. Unter Ausnutzung der Abschätzung

$$\begin{aligned}
\mathrm{E}\left(Y_1^*(t_k)^2 1_{\left\{Y_1^*(t_l)^2\geq\frac{\varepsilon^2 n}{m}\right\}}\Big|\mathcal{A}_n\right) &\leq \mathrm{E}\left(Y_1^*(t_k)^2 1_{\left\{Y_1^*(t_k)^2\geq\frac{\varepsilon^2 n}{m}\right\}}\Big|\mathcal{A}_n\right) \\
&\quad + \mathrm{E}\left(Y_1^*(t_l)^2 1_{\left\{Y_1^*(t_l)^2\geq\frac{\varepsilon^2 n}{m}\right\}}\Big|\mathcal{A}_n\right) \qquad (4.29)
\end{aligned}$$

zeigen wir für $1 \leq k \leq m$ nur

$$\mathrm{E}\left(Y_1^*(t_k)^2 1_{\left\{Y_1^*(t_k)^2\geq\frac{\varepsilon^2 n}{m}\right\}}\Big|\mathcal{A}_n\right) = o_{\mathrm{P}}(1).$$

Wir definieren

$$Z_{1n}(t_k) = 1_{\{e_1^*\leq t_k\}} - \mathrm{F}(t_k, \hat{\vartheta}_{n,res}),$$

$$Z_{2n}(t_k) = \frac{\mathrm{U}(t_k)}{\sigma^2} e_1^*$$

und

$$Z_{3n}(t_k) = \Delta(t_k, \vartheta_0)^t \mathrm{L}(e_1^*, \vartheta_0),$$

so daß also

$$Y_1^*(t_k) = Z_{1n}(t_k) - Z_{2n}(t_k) - Z_{3n}(t_k)$$

gilt. Mit dem gleichen Argument wie in (4.29) können wir sehen, daß es ausreicht, für jedes $\varepsilon' > 0$ und jedes $k \in \{1, \ldots, m\}$ zu zeigen, daß

$$\mathrm{E}\left(Z_{sn}(t_k)^2 1_{\left\{Z_{sn}(t_k)^2 \geq \varepsilon' n\right\}} \middle| \mathcal{A}_n \right) = o_{\mathrm{P}}(1) \tag{4.30}$$

für $s \in \{1, 2, 3\}$ gilt.

Für Z_{1n} ist dies klar, denn $|Z_{1n}|$ ist durch Eins beschränkt, weshalb die Menge im Indikator für $n > 1/\varepsilon'$ leer ist.

Weiter gilt im Fall $\mathrm{U}(t_k) \neq 0$ ebenfalls

$$\mathrm{E}\left(Z_{2n}(t_k)^2 1_{\left\{Z_{2n}(t_k)^2 \geq \varepsilon' n\right\}} \middle| \mathcal{A}_n \right) = \frac{\mathrm{U}(t_k)^2}{\sigma^4} \mathrm{E}\left(e_1^{*2} 1_{\left\{e_1^{*2} \geq \frac{\varepsilon' \sigma^4 n}{\mathrm{U}(t_k)^2}\right\}} \middle| \mathcal{A}_n \right) = o_{\mathrm{P}}(1)$$

gemäß (4.1) und (4.6), und im Fall $\mathrm{U}(t_k) = 0$ ist Z_{2n} trivialerweise gleich Null.

Schließlich bleibt Z_{3n} zu betrachten. Ist $\|\Delta(t_k, \vartheta_0)\| = 0$, so ist nichts zu zeigen. Andernfalls gilt

$$\begin{aligned}
&\mathrm{E}\left(Z_{3n}(t_k)^2 1_{\{Z_{3n}(t_k)^2 \geq \varepsilon' n\}} \middle| \mathcal{A}_n \right) \\
&\leq \|\Delta(t_k, \vartheta_0)\|^2 \mathrm{E}\left(\left\| \mathrm{L}(e_1^*, \hat{\vartheta}_{n,res}) \right\|^2 1_{\left\{\left\|\mathrm{L}(e_1^*, \hat{\vartheta}_{n,res})\right\|^2 \geq \frac{\varepsilon' n}{\|\Delta(t_k, \vartheta_0)\|^2}\right\}} \middle| \mathcal{A}_n \right) \\
&\leq \|\Delta(t_k, \vartheta_0)\|^2 \sum_{a=1}^{d} \sum_{b=1}^{d} \mathrm{E}\left(\mathrm{L}_a(e_1^*, \hat{\vartheta}_{n,res})^2 1_{\left\{L_b(e_1^*, \hat{\vartheta}_{n,res})^2 \geq \frac{\varepsilon' n}{d\|\Delta(t_k, \vartheta_0)\|^2}\right\}} \middle| \mathcal{A}_n \right),
\end{aligned}$$

weshalb wir im Folgenden nur die stochastische Konvergenz der einzelnen Summanden gegen Null zeigen müssen. Dies reduziert sich unter nochmaliger Verwendung des Arguments aus (4.29) zu der Forderung

$$\mathrm{E}\left(\mathrm{L}_a(e_1^*, \hat{\vartheta}_{n,res})^2 1_{\left\{L_a(e_1^*, \hat{\vartheta}_{n,res})^2 \geq \varepsilon'' n\right\}} \middle| \mathcal{A}_n \right) = o_{\mathrm{P}}(1)$$

für jedes $a \in \{1, \ldots, d\}$ und beliebiges $\varepsilon'' > 0$. Dies ist aber wegen (4.1) und (4.26) erfüllt, so daß die Behauptung (4.30) auch für Z_{3n} gilt, was den Beweis der Lindeberg-Bedingung abschließt.

Zusammen mit der oben gezeigten Normierungsbedingung impliziert der mehrdimensionale zentrale Grenzwertsatz für zeilenweise unabhängige Dreiecksschemata die Konvergenz der endlich dimensionalen Randverteilungen

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underline{Y}_i^*(\underline{t}_m) \xrightarrow{\mathcal{L}} \mathcal{N}_m(0,\Gamma) \qquad \text{unter } \mathrm{P}\left(\,\cdot\, | \mathcal{A}_n\right) \text{ P-stochastisch.}$$

Nun beweisen wir die C-Straffheit des Prozesses $(n^{-1/2}\sum_{i=1}^n Y_i^*)_{n\in\mathbb{N}}$. Also ist für beliebige $\varepsilon, \varepsilon' > 0$ zu zeigen, daß

$$\limsup_{n\to\infty} \mathrm{P}\left(\mathrm{P}\left(\mathrm{w}_\infty\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i^*, \delta\right) \geq \varepsilon' \Big| \mathcal{A}_n\right) \geq \varepsilon\right) \longrightarrow 0 \qquad \text{für } \delta \to 0$$

gilt. Wegen der Summenstruktur der Y_i^* reicht es wie schon im Beweis zu (2.22), die Straffheit der einzelnen Summanden zu beweisen. Wir untersuchen also die Prozesse P_{sn} für $s \in \{1,2,3\}$, die wie folgt definiert sind:

$$P_{1n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1_{\{e_i^*\leq\cdot\}} - \mathrm{F}(\cdot, \hat{\vartheta}_{n,res})\right),$$

$$P_{2n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\mathrm{U}}{\sigma^2} e_i^*$$

und

$$P_{3n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Delta(\cdot, \vartheta_0)^t \mathrm{L}(e_i^*, \hat{\vartheta}_{n,res}).$$

Sowohl der Prozess P_{2n} als auch P_{3n} lassen sich wegen (4.8) und (4.17) jeweils als Skalarprodukt einer deterministischen Funktion $\mathrm{h} : \mathbb{R} \longrightarrow \mathbb{R}^m$ mit $m \in \mathbb{N}$ und eines vom Argument des Prozesses unabhängigen m-dimensionalen Zufallsvektors W_n mit

$$W_n = O_{\mathrm{P}^*}(1) \qquad \text{unter } \mathrm{P}\left(\,\cdot\, | \mathcal{A}_n\right) \text{ P-stochastisch}$$

schreiben. Die Funktion h ist dabei jeweils von n unabhängig und in x gleichmäßig stetig. Für U gilt das natürlich, und $\Delta(\cdot, \vartheta_0)$ erfüllt dies wegen (4.18) und (2.7).

Da h gleichmäßig stetig ist, existiert zu $\varepsilon', C > 0$ ein $\delta_0(\varepsilon', C) > 0$, so daß für alle $0 < \delta < \delta_0(\varepsilon', C)$ die Abschätzung

$$\mathrm{w}_\infty(\mathrm{h}, \delta) < \frac{\varepsilon'}{C}$$

gilt. Folglich gilt für alle kleinen δ auch

$$\begin{aligned} \mathrm{P}\left(\mathrm{w}_\infty\left(\mathrm{h}^t W_n, \delta\right) \geq \varepsilon' | \mathcal{A}_n\right) &\leq \mathrm{P}\left(\mathrm{w}_\infty(\mathrm{h}, \delta)\|W_n\| \geq \varepsilon' | \mathcal{A}_n\right) \\ &\leq \mathrm{P}\left(\|W_n\| > C | \mathcal{A}_n\right). \end{aligned}$$

Auf Grund von $W_n = O_{\mathrm{P}^*}(1)$ unter $\mathrm{P}\left(\,\cdot\, |\mathcal{A}_n\right)$ P-stochastisch gibt es aber ein $C > 0$ mit

$$\limsup_{n\to\infty} \mathrm{P}\left(\mathrm{P}(\|W_n\| > C|\mathcal{A}_n) \geq \varepsilon\right) \leq \varepsilon,$$

und so können wir für $0 < \delta < \delta_0(\varepsilon', C)$ insgesamt

$$\limsup_{n\to\infty} \mathrm{P}\left(\mathrm{P}\left(\mathrm{w}_\infty(\mathrm{h}^t W_n, \delta) \geq e'|\mathcal{A}_n\right) \geq \varepsilon\right) \leq \varepsilon$$

folgern, was die C-Straffheit der Prozesse P_{2n} und P_{3n} beweist.

Es bleibt also P_{1n} zu diskutieren. Seien $\varepsilon, \varepsilon' > 0$ und $U_1, \ldots, U_n$ unabhängige auf [0,1] gleichverteilte Zufallsvariablen, die außerdem von $\mathcal{A}_n$ unabhängig sind. Setzen wir

$$d_i = \mathrm{F}^{-1}(U_i, \hat{\vartheta}_{n,res}) \qquad \text{für } 1 \leq i \leq n,$$

wobei $\mathrm{F}^{-1}(\cdot, \hat{\vartheta}_{n,res})$ die Quantilfunktion zu $\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})$ bezeichnet, so sind auch die d_i unter $\mathrm{P}(\,\cdot\, |\mathcal{A}_n)$ unabhängig, und es gilt

$$(d_1, \ldots, d_n) \stackrel{\mathcal{L}}{=} (e_1^*, \ldots, e_n^*)$$

unter $\mathrm{P}(\,\cdot\, |\mathcal{A}_n)$. Setzen wir weiter

$$\alpha_n(u) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n 1_{\{U_i \leq u\}} - u\right) \qquad \text{für } u \in [0,1],$$

so gilt

$$P_{1n} \stackrel{\mathcal{L}}{=} \alpha_n(\mathrm{F}(\cdot, \hat{\vartheta}_{n,res}))$$

unter $\mathrm{P}(\,\cdot\, |\mathcal{A}_n)$ und folglich

$$\mathrm{P}\left(\mathrm{w}_\infty\left(P_{1n}, \delta\right) \geq \varepsilon' |\mathcal{A}_n\right) = \mathrm{P}\left(\mathrm{w}_\infty\left(\alpha_n\left(\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})\right), \delta\right) \geq \varepsilon' \middle| \mathcal{A}_n\right)$$

für beliebiges $\delta > 0$. Auf Grund der bekannten C-Straffheit des uniformen empirischen Prozesses gibt es ein $\delta_0 > 0$ so daß

$$\limsup_{n\to\infty} \mathrm{P}\left(w_\infty(\alpha_n(\mathrm{F}(\cdot, \vartheta_0)), \delta) \geq \frac{\varepsilon'}{2}\right) \leq \frac{\varepsilon}{4} \tag{4.31}$$

für alle $0 < \delta < \delta_0$ gilt. Deshalb gilt für alle solchen δ die Abschätzung

$$\begin{aligned}
&\mathrm{P}\left(\mathrm{P}\left(\mathrm{w}_\infty\left(\alpha_n\left(\mathrm{F}(\cdot, \hat{\vartheta}_{n,res})\right), \delta\right) \geq \varepsilon' \middle| \mathcal{A}_n\right) \geq \varepsilon\right) \\
\leq\ & \mathrm{P}\left(\mathrm{P}\left(\sup_{x\in\mathbb{R}} \left|\alpha_n\left(\mathrm{F}(x, \hat{\vartheta}_{n,res})\right) - \alpha_n\left(\mathrm{F}(x, \vartheta_0)\right)\right| \geq \frac{\varepsilon'}{4} \middle| \mathcal{A}_n\right) \geq \frac{\varepsilon}{2}\right) \\
&+\mathrm{P}\left(\mathrm{P}\left(\mathrm{w}_\infty(\alpha_n(\mathrm{F}(\cdot, \vartheta_0)), \delta) \geq \frac{\varepsilon'}{2} \middle| \mathcal{A}_n\right) \geq \frac{\varepsilon}{2}\right),
\end{aligned}$$

und wegen der Unabhängigkeit der U_i von $\mathcal{A}_n$ und (4.31) ist der zweite Summand für alle großen n gleich Null. Für den ersten erhalten wir jedoch für beliebiges $\delta' > 0$ die folgende Abschätzung:

$$\begin{aligned} &\mathrm{P}\left(\mathrm{P}\left(\sup_{x\in\mathbb{R}}\left|\alpha_n\left(\mathrm{F}(x,\hat{\vartheta}_{n,res})\right)-\alpha_n\left(\mathrm{F}(x,\vartheta_0)\right)\right|\geq\frac{\varepsilon'}{4}\middle|\mathcal{A}_n\right)\geq\frac{\varepsilon}{2}\right)\\ \leq\ &\mathrm{P}\left(\mathrm{P}\left(\sup_{\substack{s,t\in[0,1]\\|s-t|\leq\delta'}}|\alpha_n(s)-\alpha_n(t)|\geq\frac{\varepsilon'}{4}\middle|\mathcal{A}_n\right)\geq\frac{\varepsilon}{4}\right)\\ &+\mathrm{P}\left(\sup_{x\in\mathbb{R}}\left|\mathrm{F}(x,\hat{\vartheta}_{n,res})-\mathrm{F}(x,\vartheta_0)\right|>\delta'\right). \end{aligned}$$

Wegen der Unabhängigkeit der U_i von $\mathcal{A}_n$ und der C-Straffheit des uniformen empirischen Prozesses ist der Limes superior der ersten Wahrscheinlichkeit für alle kleinen δ' gleich Null. Die zweite Wahrscheinlichkeit andererseits ist wegen (3.26) eine Nullfolge.

Damit ist die C-Straffheit von P_{1n} und somit auch von $(n^{-1/2}\sum_{i=1}^n Y_i^*)_{n\in\mathbb{N}}$ bewiesen. Zusammen mit der Konvergenz der endlich dimensionalen Randverteilungen beweist dies den funktionalen Grenzwertsatz (4.27). □

Zum Schluß dieses Kapitels wenden wir uns noch einmal der Frage der Berechnung kritischer Werte der Kolmogorov-Smirnov-Statistik zu. Wir haben gesehen, daß unter geeigneten Voraussetzungen die funktionalen Grenzwertsätze (3.27) und (4.27) gelten. Dabei ist es wesentlich, daß der Grenzprozeß W in beiden Fällen derselbe ist. Nun kann nämlich gezeigt werden, daß die Differenz der Quantile der Verteilungen von $\sup_{x\in\mathbb{R}}|\mathrm{F}_{n,res,z}(x)-\mathrm{F}(x,\hat{\vartheta}_{n,res})|$ und $\sup_{x\in\mathbb{R}}|\mathrm{F}^*_{n,z}(x)-\mathrm{F}(x,\hat{\vartheta}^*_n)|$ stochastisch gegen Null konvergiert. Ist

$$\mathrm{H}_n^*(t)=\mathrm{P}\left(\sup_{x\in\mathbb{R}}\left|\mathrm{F}^*_{n,z}(x)-\mathrm{F}(x,\hat{\vartheta}^*_n)\right|\leq t\middle|\mathcal{A}_n\right)\qquad\text{für } t\in\mathbb{R}$$

und H_n^{*-1} die Quantilfunktion zu H_n^*, so erhalten wir zu $\alpha\in(0,1)$ mit $\mathrm{H}_n^{*-1}(1-\alpha)$ einen asymptotischen kritischen Wert für die Kolmogorov-Smirnov-Statistik, das heißt es gilt

$$\mathrm{P}\left(\sup_{x\in\mathbb{R}}\left|\mathrm{F}_{n,res,z}(x)-\mathrm{F}(x,\hat{\vartheta}_{n,res})\right|\geq\mathrm{H}_n^{*-1}(1-\alpha)\right)\longrightarrow\alpha\qquad\text{für } n\to\infty;$$

für die technische Durchführung dieser Argumente siehe z.B. Abschnitt 4.2.1 in Shao und Tu (1995).

Die Verteilung von H_n^* und damit auch das Quantil $\mathrm{H}_n^{*-1}(1-\alpha)$ sind prinzipiell bekannt. In der Praxis kann man $\mathrm{H}_n^{*-1}(1-\alpha)$ allerdings meist nicht oder nur mit sehr großem Aufwand ausrechnen, so daß man dieses Quantil mittels einer Simulation approximiert.

Kapitel 5

Nichtparametrische Ein-Schritt-Prognoseintervalle in stabilen autoregressiven Prozessen erster Ordnung

5.1 Grundlagen und technische Hilfsmittel

Wir betrachten im Folgenden einen stabilen AR(1)-Prozeß $(X_n)_{n\in\mathbb{N}}$ mit Fehlerfolge $(e_i)_{i\in\mathbb{N}}$ und Autoregressionsparameter ρ. Die e_i sind wie bisher unabhängig und identisch nach der stetigen Verteilungsfunktion F verteilt. Weiter setzen wir $X_0 = 0$ und

$$\mathrm{E}\left(e_1\right) = 0 \qquad \text{und} \qquad 0 < \sigma^2 := \mathrm{Var}\left(e_1\right) = \mathrm{E}\left(e_1^2\right) < \infty \tag{2.10}$$

sowie

$$\text{F hat eine gleichmäßig stetige, } \lambda - f\ddot{u} \text{ positive Dichte f,} \tag{3.13}$$

voraus. Wenn wir den Prozeß bis zum Zeitpunkt n beobachtet haben, so stellt sich natürlicherweise die Frage nach einer Prognose für X_{n+1} auf der Basis der beobachteten $X_0, \ldots, X_n$. Wegen

$$X_{n+1} = \rho X_n + e_{n+1}$$

und $\mathrm{E}(e_{n+1}) = \mathrm{E}(e_1) = 0$ ist $\hat{\rho}_n X_n$ mit einem konsistenten Schätzer $\hat{\rho}_n$ für ρ die natürliche *Prognose* für X_{n+1}. Bezeichnen wir mit H^{-1} die Quantilfunktion zu einer Verteilungsfunktion H, so sind wegen

$$\begin{aligned} &\mathrm{P}\left(\rho X_n + \mathrm{F}^{-1}\left(\frac{\alpha}{2}\right) \leq X_{n+1} \leq \rho X_n + \mathrm{F}^{-1}\left(1-\frac{\alpha}{2}\right) \middle| \mathcal{A}_n\right) \\ = \;& \mathrm{P}\left(\mathrm{F}^{-1}\left(\frac{\alpha}{2}\right) \leq e_{n+1} \leq \mathrm{F}^{-1}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha \end{aligned}$$

die Intervalle

$$\hat{I}_n = \left[\, \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}\left(\frac{\alpha}{2}\right), \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}\left(1 - \frac{\alpha}{2}\right) \, \right] \tag{5.1}$$

mit $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ oder $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ aus Kapitel 3 natürliche nichtparametrische (Plug-In-)*Prognoseintervalle* für X_{n+1} zu einem fest vorgegebenen Niveau $\alpha \in (0,1)$.

Der wohl bekannteste Schätzer für ρ ist der durch

$$\hat{\rho}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2} = \rho + \frac{\sum_{i=1}^n X_{i-1} e_i}{\sum_{i=1}^n X_{i-1}^2} \tag{5.2}$$

definierte Kleinst-Quadrate-Schätzer (KQS), auf den wir uns im Folgenden beschränken werden.

Die mit Hilfe dieses Schätzers gemäß (5.1) konstruierten Intervalle sind konsistent im Sinne von

$$\mathrm{P}\left(X_{n+1} \in \hat{I}_n | \mathcal{A}_n\right) \longrightarrow 1 - \alpha \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.} \tag{5.3}$$

Dies zu zeigen ist ein erstes Ziel dieses Kapitels. Darüber hinaus werden wir sogar zeigen, daß die Zufallsfolge

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in \hat{I}_n | \mathcal{A}_n\right) - (1 - \alpha)\right)$$

verteilungskonvergent ist und ihre Grenzverteilung angeben.

Dazu benötigen wir einige technische Hilfsmittel.

5.1.1 Das asymptotische Verhalten von Quantilen

Wie wir an der Definition (5.1) sehen, sind Quantile der Verteilungsschätzer $\mathrm{F}_{n,res}$ und $\mathrm{F}_{n,res,z}$ von großer Bedeutung für die Konstruktion nichtparametrischer Prognoseintervalle. Deshalb ist es sinnvoll, das asymptotische Verhalten von Quantilen zu studieren. Das folgende technische Lemma hat einen elementaren Beweis, den wir nicht ausführen. Für die Beweismethoden vgl. zum Beispiel Witting und Noelle (1970), Satz 2.11.

Lemma 5.1

Sei F eine stetige Verteilungsfunktion und F_n ein beliebiger Schätzer für F, das heißt eine Folge von zufälligen Verteilungsfunktionen, mit

$$\sup_{x \in \mathbb{R}} |\mathrm{F}_n(x) - \mathrm{F}(x)| = o_{\mathrm{P}}(1)$$

und seien weiter F_n^{-1} und F^{-1} die zugehörigen Quantilfunktionen. Dann gilt für alle $u \in (0,1)$, in denen F^{-1} stetig ist, die Aussage

$$\mathrm{F}_n^{-1}(u) \longrightarrow \mathrm{F}^{-1}(u) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.} \tag{5.4}$$

Unser nächstes Ziel ist es, die Differenzen $\mathrm{F}_n^{-1}(u) - \mathrm{F}^{-1}(u)$ und $\mathrm{F}_n(\mathrm{F}^{-1}(u)) - u$ im Sinne des Satzes von Bahadur zu vergleichen. Dazu verallgemeinern wir das Hauptresultat aus Ghosh (1971). Als Hilfsaussage benötigen wir dazu das folgende Lemma.

Lemma 5.2 *(vgl. Lemma 1 aus Ghosh (1971))*

Seien $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ Folgen von Zufallsvariablen mit den folgenden Eigenschaften:

(i) Es ist
$$X_n = O_{\mathrm{P}}(1),$$

(ii) für alle $r \in \mathbb{R}$ und alle $\xi > 0$ gilt
$$\mathrm{P}(X_n \leq r,\ Y_n \geq r + \xi) \longrightarrow 0 \qquad \text{für } n \to \infty$$
und
$$\mathrm{P}(Y_n \leq r,\ X_n \geq r + \xi) \longrightarrow 0 \qquad \text{für } n \to \infty.$$

Dann gilt
$$X_n - Y_n \longrightarrow 0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Dieses Lemma spielt eine zentrale Rolle für den Beweis des folgenden Satzes vom Bahadur-Typ. Zur Abkürzung benutzen wir hier und im Folgenden zu $x \in \mathbb{R}$ und $\delta > 0$ die Bezeichnung

$$V_\delta(x) = \{y \in \mathbb{R} : |x - y| \leq \delta\}.$$

Satz 5.3

Sei $u \in (0,1)$ und F eine Verteilungsfunktion, die an der Stelle $\mathrm{F}^{-1}(u)$ differenzierbar ist mit der Ableitung
$$\mathrm{f}(\mathrm{F}^{-1}(u)) > 0.$$

Sei F_n ein Schätzer für F mit

$$\sqrt{n}\left(\mathrm{F}_n(\mathrm{F}^{-1}(u)) - u\right) = O_{\mathrm{P}}(1) \tag{5.5}$$

und

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P}\left(\sup_{t \in V_\delta(\mathrm{F}^{-1}(u))} \sqrt{n} \left| \mathrm{F}_n(t) - \mathrm{F}(t) - \mathrm{F}_n(\mathrm{F}^{-1}(u)) + u \right| \geq \varepsilon \right) = 0 \qquad (5.6)$$

für jedes $\varepsilon > 0$. Dann gilt

$$\mathrm{f}(\mathrm{F}^{-1}(u))(\mathrm{F}_n^{-1}(u) - \mathrm{F}^{-1}(u)) = -(\mathrm{F}_n(\mathrm{F}^{-1}(u)) - u) + o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Beweis Wir setzen

$$X_n = \sqrt{n}\, \frac{u - \mathrm{F}_n(\mathrm{F}^{-1}(u))}{\mathrm{f}(\mathrm{F}^{-1}(u))} \qquad \text{und} \qquad Y_n = \sqrt{n}\left(\mathrm{F}_n^{-1}(u) - \mathrm{F}^{-1}(u)\right),$$

und die Behauptung entspricht gerade

$$X_n - Y_n \longrightarrow 0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Dies zeigen wir mit Lemma 5.2. Es ist nach Voraussetzung $X_n = O_{\mathrm{P}}(1)$. Sei nun $r \in \mathbb{R}$ und $\xi > 0$ beliebig sowie

$$r_n = \sqrt{n}\, \frac{\mathrm{F}(\mathrm{F}^{-1}(u) + r/\sqrt{n}) - \mathrm{F}(\mathrm{F}^{-1}(u))}{\mathrm{f}(\mathrm{F}^{-1}(u))}.$$

Aufgrund der Differenzierbarkeit von F im Punkt $\mathrm{F}^{-1}(u)$ gilt

$$r_n \longrightarrow r \qquad \text{für } n \to \infty,$$

und wir erhalten

$$\begin{aligned} \sqrt{n}\left(\mathrm{F}_n^{-1}(u) - \mathrm{F}^{-1}(u)\right) \leq r &\Leftrightarrow \mathrm{F}_n(\mathrm{F}^{-1}(u) + r/\sqrt{n}) \geq u \\ \Leftrightarrow \quad \sqrt{n}\, &\frac{\mathrm{F}(\mathrm{F}^{-1}(u) + r/\sqrt{n}) - \mathrm{F}_n(\mathrm{F}^{-1}(u) + r/\sqrt{n})}{\mathrm{f}(\mathrm{F}^{-1}(u))} \leq r_n. \end{aligned}$$

Nun gibt es zu beliebigem $\delta > 0$ ein $n_0(\delta, \xi) \in \mathbb{N}$, so daß für alle $n \geq n_0(\delta, \xi)$ die Abschätzung

$$\begin{aligned} &\mathrm{P}(Y_n \leq r,\ X_n \geq r + \xi) \\ &= \mathrm{P}\left(\sqrt{n}\, \frac{\mathrm{F}(\mathrm{F}^{-1}(u) + r/\sqrt{n}) - \mathrm{F}_n(\mathrm{F}^{-1}(u) + r/\sqrt{n})}{\mathrm{f}(\mathrm{F}^{-1}(u))} \leq r_n,\ X_n \geq r + \xi\right) \\ &\leq \mathrm{P}\left(\sup_{t \in V_\delta(\mathrm{F}^{-1}(u))} \sqrt{n} \left| \mathrm{F}_n(t) - \mathrm{F}(t) - \mathrm{F}_n(\mathrm{F}^{-1}(u)) + u \right| \geq \frac{\xi}{2} \mathrm{f}(\mathrm{F}^{-1}(u))\right) \end{aligned}$$

gilt, und der Limes Superior der letzten Wahrscheinlichkeit konvergiert für $\delta \downarrow 0$ wegen der Voraussetzung (5.6) gegen Null.

Definieren wir andererseits $r' = r + \xi/2$ und r'_n analog zu r_n, so gilt für große n die Ungleichung

$$\begin{aligned}
&\mathrm{P}(X_n \leq r,\ Y_n \geq r+\xi) \\
&\quad\leq\ \mathrm{P}\left(X_n \leq r' - \frac{\xi}{2},\ Y_n > r'\right) \\
&\quad=\ \mathrm{P}\left(X_n \leq r' - \frac{\xi}{2},\ \sqrt{n}\ \frac{\mathrm{F}(\mathrm{F}^{-1}(u) + r'/\sqrt{n}) - \mathrm{F}_n(\mathrm{F}^{-1}(u) + r'/\sqrt{n})}{\mathrm{f}(\mathrm{F}^{-1}(u))} > r'_n\right) \\
&\quad\leq\ \mathrm{P}\left(\sup_{t \in V_\delta(\mathrm{F}^{-1}(u))} \sqrt{n}\left|\mathrm{F}_n(t) - \mathrm{F}(t) - \mathrm{F}_n(\mathrm{F}^{-1}(u)) + u\right| \geq \frac{\xi}{4}\mathrm{f}(\mathrm{F}^{-1}(u))\right),
\end{aligned}$$

was ebenfalls mit (5.6) die gewünschte Konvergenz impliziert.

Insgesamt folgt aus dem Bewiesenen mit Hilfe von Lemma 5.2 die Behauptung. □

In Ghosh (1971) wird Satz 5.3 nur für die empirische Verteilungsfunktion F_n zu unabhängig und identisch verteilten Zufallsvariablen formuliert und bewiesen, wobei natürlich (5.5) und (5.6) automatisch erfüllt sind.

Ist die Folge

$$\sqrt{n}\left(\mathrm{F}_n(\mathrm{F}^{-1}(u)) - u\right)$$

verteilungskonvergent, so impliziert der gerade bewiesene Satz auch die Verteilungskonvergenz von

$$\sqrt{n}(\mathrm{F}_n^{-1}(u) - \mathrm{F}^{-1}(u)).$$

Für die Voraussetzung (5.6) ist insbesondere die C-Straffheit des stochastischen Prozesses $(\sqrt{n}(\mathrm{F}_n - \mathrm{F}))_{n\in\mathbb{N}}$ hinreichend.

5.1.2 Der mehrdimensionale zentrale Grenzwertsatz für Martingaldifferenzschemata

Wir betrachten an dieser Stelle den Kleinst-Quadrate-Schätzer $\hat{\rho}_n$, den wir bei der Konstruktion von $\hat{I}_n$ verwenden, beziehungsweise die zu (5.2) äquivalente Darstellung

$$\sqrt{n}(\hat{\rho}_n - \rho) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_{i-1}e_i}{\frac{1}{n}\sum_{i=1}^n X_{i-1}^2}. \tag{5.7}$$

Das asymptotische Verhalten von $n^{1/2}(\hat{\rho}_n - \rho)$ ist wohlbekannt. Es gilt nämlich

$$\frac{1}{n}\sum_{i=1}^n X_{i-1}^2 \longrightarrow \frac{\sigma^2}{1-\rho^2} \qquad \text{für } n \to \infty \qquad \text{P-stochastisch} \tag{5.8}$$

und

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} e_i \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^4}{1-\rho^2}\right) \qquad \text{für } n \to \infty, \tag{5.9}$$

also mit dem Satz von Cramér-Slutzky

$$\sqrt{n}\,(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1-\rho^2) \qquad \text{für } n \to \infty, \tag{5.10}$$

insbesondere ist für den KQS in stabilen AR(1)-Modellen stets (3.2) erfüllt. Der Beweis von (5.8) kann mit elementaren Methoden geführt werden. Eine Möglichkeit zum Beweis von (5.9) ist wegen der Martingalstruktur des Zählers von (5.7) der *zentrale Grenzwertsatz für Martingaldifferenzschemata,* den wir gleich in einer mehrdimensionalen Version angeben werden. Dieser mehrdimensionale zentrale Grenzwertsatz für Martingaldifferenzschemata wird ein wesentliches Hilfsmittel in späteren Beweisen sein.

Wir betrachten Dreiecksschemata $(\underline{Y}_{nk}, \mathcal{F}_{nk})_{1 \le k \le k_n}$ mit $k_n \in \mathbb{N}$ von d-dimensionalen Zufallsvektoren $\underline{Y}_{nk}$ und Sub-σ-Algebren $\mathcal{F}_{nk}$ von $\mathcal{A}$. Ein solches Dreiecksschema $(\underline{Y}_{nk}, \mathcal{F}_{nk})_{1 \le k \le k_n}$ heißt *quadratisch integrierbares Martingaldifferenzschema* (MDS), falls es für jedes $n \in \mathbb{N}$ eine d-dimensionale Martingaldifferenzfolge ist, das heißt wenn für beliebiges $n \in \mathbb{N}$ die folgenden Bedingungen gelten:

(i) $(\mathcal{F}_{nk})_{1 \le k \le k_n}$ ist eine Filtration,

(ii) $\mathrm{Var}(\underline{Y}_{nk})$ existiert für $1 \le k \le k_n$,

(iii) $\underline{Y}_{nk}$ ist bezüglich $\mathcal{F}_{nk}$ meßbar für $1 \le k \le k_n$,

(iv) $\mathrm{E}(\underline{Y}_{nk} | \mathcal{F}_{n,k-1}) = 0$ für $1 \le k \le k_n$ mit $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Einfaches Nachrechnen zeigt, daß diese Bedingungen für $(Y_{ni}, \mathcal{A}_i)_{1 \le i \le n}$ mit

$$Y_{ni} = \frac{X_{i-1} e_i}{\sqrt{n}}$$

erfüllt sind. Auch jede Folge unabhängig und identisch verteilter, zentrierter und quadratisch integrierbarer Zufallsvariablen bildet ein MDS bezüglich der kanonischen Filtration.

Für solche d-dimensionale quadratisch integrierbare MDS gilt der folgende zentrale Grenzwertsatz.

Satz 5.4

Sei $(\underline{Y}_{nk}, \mathcal{F}_{nk})_{1\le k\le k_n}$ ein quadratisch integrierbares d-dimensionales MDS und es gelte die konditionierte Normierungsbedingung

$$\sum_{k=1}^{k_n} \text{Var}\left(\underline{Y}_{nk}^2|\mathcal{F}_{n,k-1}\right) \longrightarrow \Gamma \qquad \text{für } n\to\infty \qquad \text{P-stochastisch}$$

mit einer nichtnegativ definiten Matrix Γ. Weiter gelte für jede Komponente der Vektoren des MDS die konditionierte Lindebergbedingung, das heißt für alle $\varepsilon > 0$ und $1 \le l \le d$ gilt

$$\sum_{k=1}^{k_n} \text{E}\left((\underline{Y}_{nk})_l^2 1_{\{|(\underline{Y}_{nk})_l|\ge\varepsilon\}}|\mathcal{F}_{n,k-1}\right) \longrightarrow 0 \qquad \text{für } n\to\infty \qquad \text{P-stochastisch,}$$

dann gilt

$$\sum_{k=1}^{k_n} \underline{Y}_{nk} \xrightarrow{\mathcal{L}} \mathcal{N}_d(0,\Gamma) \qquad \text{für } n\to\infty.$$

Beweis Mit zu (4.29) analogen Argumenten impliziert die komponentenweise konditionierte Lindebergbedingung für jedes $\varepsilon > 0$ die konditionierte Lindebergbedingung

$$\sum_{k=1}^{k_n} \text{E}\left(\|\underline{Y}_{nk}\|^2\, 1_{\{\|\underline{Y}_{nk}\|\ge\varepsilon\}}|\mathcal{F}_{n,k-1}\right) \longrightarrow 0 \qquad \text{für } n\to\infty \qquad \text{P-stochastisch,}$$

und der Beweis des Satzes folgt mit dem Cramér-Wold-Device aus der eindimensionalen Version des Satzes. □

Als Beispiel betrachten wir nochmals das MDS $(Y_{ni}, \text{A}_i)_{1\le i\le n}$ mit

$$Y_{ni} = \frac{X_{i-1}e_i}{\sqrt{n}}.$$

Nachrechnen der konditionierten Normierungsbedingung und der konditionierten Lindebergbedingung ergibt dann die Verteilungskonvergenzaussage (5.9).

5.2 Das asymptotische Verhalten nichtparametrischer Prognoseintervalle

Wir sind nun in der Lage, die bereits in (5.3) behauptete Konsistenz der gemäß (5.1) definierten Prognoseintervalle

$$\hat{I}_n = \left[\ \hat{\rho}_n X_n + \hat{\text{F}}_n^{-1}\left(\frac{\alpha}{2}\right), \hat{\rho}_n X_n + \hat{\text{F}}_n^{-1}\left(1-\frac{\alpha}{2}\right)\ \right]$$

mit $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ oder $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ zu beweisen. Zur Notationsvereinfachung setzen wir zu beliebigem $\alpha \in (0,1)$ im Folgenden

$$v = \frac{\alpha}{2} \qquad \text{und} \qquad w = 1 - \frac{\alpha}{2}.$$

Dabei gilt offensichtlich $0 < v < 1/2 < w < 1$.

Lemma 5.5

Unter den Voraussetzungen (2.10) und (3.13) gilt

$$\mathrm{P}\left(X_{n+1} \in \hat{I}_n | \mathcal{A}_n\right) \longrightarrow 1 - \alpha \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.} \tag{5.3}$$

Beweis Es gilt

$$\begin{aligned} &\mathrm{P}\left(X_{n+1} \in \left[\ \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(v), \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(w)\ \right] \Big| \mathcal{A}_n\right) \\ &= \ \mathrm{P}\left((\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(v) \leq e_{n+1} \leq (\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(w) \Big| \mathcal{A}_n\right) \\ &= \ \mathrm{F}\left((\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(w)\right) - \mathrm{F}\left((\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(v)\right), \end{aligned}$$

da F stetig ist. Wir erinnern noch einmal daran, daß $\hat{\rho}_n$ in diesem und dem folgenden Kapitel stets den KQS bezeichnet, und daß deshalb

$$\sqrt{n}\,(\hat{\rho}_n - \rho) = O_{\mathrm{P}}(1) \tag{5.11}$$

gilt. Außerdem sehen wir leicht

$$X_n = \sum_{i=1}^{n} \rho^{n-i} e_i \stackrel{\mathcal{L}}{=} \sum_{i=1}^{n} \rho^{i-1} e_i = Y_n,$$

und Y_n konvergiert fast sicher und in L_2 gegen

$$Y_\infty = \sum_{i=1}^{\infty} \rho^{i-1} e_i.$$

Also gilt auch

$$X_n \xrightarrow{\mathcal{L}} Y_\infty \qquad \text{für } n \to \infty, \tag{5.12}$$

und insbesondere folgt

$$X_n = O_{\mathrm{P}}(1).$$

Das Produkt $(\hat{\rho}_n - \rho)X_n$ konvergiert somit stochastisch gegen Null.

Andererseits stellen die Voraussetzungen (2.10), (3.13) und (5.11) die Gültigkeit von (3.14) sowie (3.18) sicher, was insbesondere

$$\sup_{x\in\mathbb{R}} |\hat{\mathrm{F}}_n(x) - \mathrm{F}(x)| = o_{\mathrm{P}}(1)$$

impliziert. Wegen (3.13) ist F^{-1} auf $(0,1)$ stetig, so daß wir mit Lemma 5.1 für alle $u \in (0,1)$ die stochastische Konvergenz

$$\hat{\mathrm{F}}_n^{-1}(u) \longrightarrow \mathrm{F}^{-1}(u) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch}$$

folgern können.

Insgesamt ergibt sich

$$(\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(w) \longrightarrow \mathrm{F}^{-1}(w) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch}$$

sowie

$$(\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(v) \longrightarrow \mathrm{F}^{-1}(v) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch,}$$

und die Stetigkeit von F impliziert die Behauptung. □

Wir sehen, daß die gerade bewiesene Konsistenz für beide Verteilungsschätzer, also $\mathrm{F}_{n,res}$ und $\mathrm{F}_{n,res,z}$, beziehungsweise die aus ihnen abgeleiteten Prognoseintervalle gilt. Auch die Länge $\hat{\mathrm{F}}_n^{-1}(w) - \hat{\mathrm{F}}_n^{-1}(v)$ der Intervalle ist asymptotisch gleich, nämlich gleich $\mathrm{F}^{-1}(w) - \mathrm{F}^{-1}(v)$. Um den Effekt der unterschiedlichen Schätzung von F zu sehen, betrachten wir Folge

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in \hat{I}_n \middle| \mathcal{A}_n\right) - (1-\alpha)\right)$$

von Zufallsvariablen. Wie wir sehen werden ist diese verteilungskonvergent, wobei sich die Grenzverteilungen in den Fällen $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ und $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ unterscheiden. Wir werden zeigen, daß die Grenzvarianz im Fall $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ nichtgrößer der im Fall $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ ist, was als „schnellere“ oder „präzisere“ Konvergenz gegen das nominelle Niveau interpretiert werden kann. Ein wesentlicher Schritt zum Beweis der Verteilungskonvergenz ist die folgende stochastische Entwicklung.

Lemma 5.6

Unter den Voraussetzungen (2.10) und (3.13) gilt

$$\begin{aligned}
&\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in \left[\ \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(v), \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(w)\ \right] \middle| \mathcal{A}_n\right) - (1-\alpha)\right) \\
&\quad = \left(\mathrm{f}\left(\mathrm{F}^{-1}(w)\right) - \mathrm{f}\left(\mathrm{F}^{-1}(v)\right)\right) \frac{1-\rho^2}{\sigma^2} Z_n X_n \\
&\qquad - \begin{cases} (C_{1n}(w) - C_{1n}(v)) + o_{\mathrm{P}}(1) & \text{für } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res} \\ (C_{2n}(w) - C_{2n}(v)) + o_{\mathrm{P}}(1) & \text{für } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z} \end{cases}
\end{aligned} \tag{5.13}$$

mit

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i-1} e_i, \tag{5.14}$$

$$C_{1n}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1_{\{e_i \leq \mathrm{F}^{-1}(u)\}} - u\right) \qquad \text{für } u \in (0,1)$$

und

$$C_{2n}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1_{\{e_i \leq \mathrm{F}^{-1}(u)\}} - u - \frac{\mathrm{U}(\mathrm{F}^{-1}(u))}{\sigma^2} e_i\right) \qquad \text{für } u \in (0,1).$$

Beweis Es ist

$$\begin{aligned}
&\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in \left[\ \hat\rho_n X_n + \hat{\mathrm{F}}_n^{-1}(v), \hat\rho_n X_n + \hat{\mathrm{F}}_n^{-1}(w)\ \right] \Big| \mathcal{A}_n\right) - (1-\alpha)\right) \\
&= \sqrt{n}\left[\mathrm{F}\left((\hat\rho_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(w)\right) - \mathrm{F}(\mathrm{F}^{-1}(w)) \right.\\
&\quad \left. -\mathrm{F}\left((\hat\rho_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(v)\right) + \mathrm{F}(\mathrm{F}^{-1}(v))\right] \\
&= \sqrt{n}(S_n(w) - S_n(v)),
\end{aligned}$$

und für jedes $u \in (0,1)$ können wir aus (3.13) folgern, daß

$$S_n(u) = \left((\hat\rho_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(u) - \mathrm{F}^{-1}(u)\right)\left(\mathrm{f}(\mathrm{F}^{-1}(u)) + o_{\mathrm{P}}(1)\right)$$

gilt. Dabei ist der erste Faktor von der Ordnung $O_{\mathrm{P}}(n^{-1/2})$. Für $(\hat\rho_n - \rho)X_n$ gilt dies wegen (5.11) und (5.12). Es bleibt also

$$\hat{\mathrm{F}}_n^{-1}(u) - \mathrm{F}^{-1}(u) = O_{\mathrm{P}}\left(n^{-1/2}\right)$$

zu zeigen. Dazu betrachten wir die stochastischen Prozesse $(n^{1/2}(\mathrm{F}_{n,res} - \mathrm{F}))_{n\in\mathbb{N}}$ und $(n^{1/2}(\mathrm{F}_{n,res,z} - \mathrm{F}))_{n\in\mathbb{N}}$. Auf Grund von (3.14) und (3.18) sind sie asymptotisch äquivalent zu den Prozessen $(n^{1/2}(\mathrm{F}_n - \mathrm{F}))_{n\in\mathbb{N}}$ beziehungsweise $(n^{1/2}(\mathrm{F}_{n,z} - \mathrm{F}))_{n\in\mathbb{N}}$. Daß der empirische Prozeß $(n^{1/2}(\mathrm{F}_n - \mathrm{F}))_{n\in\mathbb{N}}$ C-straff ist und daß alle seine endlichdimensionalen Randverteilungen konvergieren, ist wohlbekannt. Daß Gleiches auch für $(n^{1/2}(\mathrm{F}_{n,z} - \mathrm{F}))_{n\in\mathbb{N}}$ gilt, wurde ebenfalls bereits gezeigt (vgl. Abschnitt 2.2.3 in Genz (2001)). Also sind die Voraussetzungen des Satzes 5.3 vom Bahadur-Typ für die Verteilungsschätzer $\mathrm{F}_{n,res}$ und $\mathrm{F}_{n,res,z}$ erfüllt, weshalb wir

$$\sqrt{n}\left(\hat{\mathrm{F}}_n^{-1}(u) - \mathrm{F}^{-1}(u)\right) = -\frac{\sqrt{n}(\hat{\mathrm{F}}_n(\mathrm{F}^{-1}(u)) - u)}{\mathrm{f}(\mathrm{F}^{-1}(u))} + o_{\mathrm{P}}(1)$$

folgern können. Die rechte Seite dieser Gleichung ist aber offensichtlich gemäß (3.14) beziehungsweise (3.18) von der Ordnung $O_{\mathrm{P}}(1)$.

Also gilt

$$S_n(u) = \Big((\hat{\rho}_n - \rho)X_n + \hat{\mathrm{F}}_n^{-1}(u) - \mathrm{F}^{-1}(u)\Big)\, \mathrm{f}(\mathrm{F}^{-1}(u)) + o_{\mathrm{P}}(n^{-1/2})$$

und somit

$$\begin{aligned}\sqrt{n}(S_n(w) - S_n(v)) \;&=\; \sqrt{n}(\hat{\rho}_n - \rho)X_n \left(\mathrm{f}(\mathrm{F}^{-1}(w)) - \mathrm{f}(\mathrm{F}^{-1}(w))\right) \\ &\quad -\sqrt{n}\left(\hat{\mathrm{F}}_n(\mathrm{F}^{-1}(w)) - w\right) + \sqrt{n}\left(\hat{\mathrm{F}}_n(\mathrm{F}^{-1}(v)) - v\right) + o_{\mathrm{P}}(1),\end{aligned}$$

wobei wir nochmals den Satz von Bahadur anwenden.

Unter Berücksichtigung der Darstellung (5.7) und der Konvergenzen (5.8) und (5.9) gilt

$$\sqrt{n}(\hat{\rho}_n - \rho) = Z_n \frac{1-\rho^2}{\sigma^2} + o_{\mathrm{P}}(1),$$

andererseits implizieren (3.14) beziehungsweise (3.18) die Darstellung

$$\sqrt{n}\left(\hat{\mathrm{F}}_n(\mathrm{F}^{-1}(u)) - u\right) = \begin{cases} C_{1n}(u) + o_{\mathrm{P}}(1) & \text{für } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res} \\ C_{2n}(u) + o_{\mathrm{P}}(1) & \text{für } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z} \end{cases}$$

für beliebiges $u \in (0,1)$, so daß der Satz vollständig bewiesen ist. □

Diese stochastische Entwicklung von

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in \left[\, \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(v), \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(w) \,\right] \Big| \mathcal{A}_n\right) - (1-\alpha)\right)$$

wird die Grundlage für eine Verteilungskonvergenzaussage bilden. Dazu untersuchen wir im nächsten Lemma zunächst das gemeinsame asymptotische Verteilungsverhalten der Folgen $Z_n, C_{1n}(v)$ und $C_{1n}(w)$.

Lemma 5.7

Unter der Voraussetztung (2.10) gilt

$$\begin{pmatrix} Z_n \\ C_{1n}(v) \\ C_{1n}(w) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_3\left(0, \Gamma_1\right) \qquad \text{für } n \to \infty \tag{5.15}$$

mit

$$\Gamma_1 = \begin{pmatrix} \frac{\sigma^4}{1-\rho^2} & 0 & 0 \\ 0 & vw & v^2 \\ 0 & v^2 & vw \end{pmatrix}.$$

Beweis Wir setzen

$$X_{ni} = \begin{pmatrix} Z_{ni} \\ C_{1ni}(v) \\ C_{1ni}(w) \end{pmatrix}$$

mit

$$Z_{ni} = \frac{X_{i-1}e_i}{\sqrt{n}}$$

und

$$C_{1ni}(u) = \frac{1_{\{e_i \leq \mathrm{F}^{-1}(u)\}} - u}{\sqrt{n}} \qquad u \in (0,1),$$

und die Behauptung ist gerade

$$\sum_{i=1}^{n} X_{ni} \xrightarrow{\mathcal{L}} \mathcal{N}_3\left(0, \Gamma_1\right) \qquad \text{für } n \to \infty.$$

Wie man leicht nachrechnet ist $(X_{ni}, \mathcal{A}_i)_{1\leq i\leq n}$ ein dreidimensionales quadratisch integrierbares MDS, so daß wir den mehrdimensionalen zentralen Grenzwertsatz für MDS anwenden können, wenn wir die konditionierte Normierungsbedingung und die komponentenweise konditionierte Lindebergbedingung nachweisen.

Zum Beweis der konditionierten Normierungsbedingung stellen wir zunächst fest, daß wegen der Unabhängigkeit von e_i und $\mathcal{A}_{i-1}$ die Gleichung

$$\sum_{i=1}^{n} \mathrm{E}\left(Z_{ni}^2 | \mathcal{A}_{i-1}\right) = \sigma^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1}^2$$

gilt, und die rechte Seite konvergiert wegen (5.8) stochastisch gegen $\sigma^4/(1-\rho^2)$. Es gilt ebenso

$$\sum_{i=1}^{n} \mathrm{E}\left(Z_{ni} C_{1ni}(u) | \mathcal{A}_{i-1}\right) = \mathrm{U}(\mathrm{F}^{-1}(u)) \frac{1}{n} \sum_{i=1}^{n} X_{i-1} = o_{\mathrm{P}}(1)$$

für jedes $u \in (0,1)$ wegen (3.5). Weiter ist

$$\sum_{i=1}^{n} \mathrm{E}\left(C_{1ni}(v) C_{1ni}(w) | \mathcal{A}_{i-1}\right) = \mathrm{E}\left(1_{\{e_1 \leq \mathrm{F}^{-1}(v)\}} 1_{\{e_1 \leq \mathrm{F}^{-1}(w)\}}\right) - vw = v - vw = v^2,$$

da stets $v < w$ gilt, und schließlich berechnen wir genauso

$$\sum_{i=1}^{n} \mathrm{E}\left(C_{1ni}^2(u) | \mathcal{A}_{i-1}\right) = u - u^2 = u(1-u) = vw$$

für $u \in \{v, w\}$. Insgesamt beweist dies die konditionierte Normierungsbedingung.

Es bleibt für jede Komponente des Vektors X_{ni} die konditionierte Lindebergbedingung zu zeigen. Für die erste Komponente ist dies wohlbekannt und wird zum Beweis der Verteilungskonvergenz (5.9) verwendet. Es genügt also die konditionierte Lindebergbedingung für die zweite und dritte Komponente zu zeigen. Sei dazu $u \in (0,1)$ und $\varepsilon > 0$. Dann gilt

$$\sum_{i=1}^{n} \mathrm{E}\left(C_{1ni}^2 1_{\{|C_{1ni}(u)| \geq \varepsilon\}} | \mathcal{A}_{i-1}\right) = \mathrm{E}\left(\left(1_{\{e_1 \leq \mathrm{F}^{-1}(u)\}} - u\right)^2 1_{\{|1_{\{e_1 \leq \mathrm{F}^{-1}(u)\}} - u| \geq \varepsilon\sqrt{n}\}}\right),$$

und der Integrand auf der rechten Seite ist wegen

$$|1_{\{e_1 \leq \mathrm{F}^{-1}(u)\}} - u| \leq 1$$

für alle großen n gleich Null, was die konditionierte Lindebergbedingung auch für die zweite und dritte Komponente von X_{ni} impliziert.

Die Anwendung des mehrdimensionalen zentralen Grenzwertsatzes für MDS schließt den Beweis des Lemmas ab. □

Eine entsprechende Verteilungskonvergenz gilt für die Folge $(Z_n, C_{2n}(v), C_{2n}(w))^t$.

Lemma 5.8

Unter der Voraussetzung (2.10) gilt

$$\begin{pmatrix} Z_n \\ C_{2n}(v) \\ C_{2n}(w) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_3\left(0, \Gamma_2\right) \qquad \text{für } n \to \infty \tag{5.16}$$

mit

$$\Gamma_2 = \begin{pmatrix} \frac{\sigma^4}{1-\rho^2} & 0 & 0 \\ 0 & vw - \frac{\mathrm{U}(\mathrm{F}^{-1}(v))^2}{\sigma^2} & v^2 - \frac{\mathrm{U}(\mathrm{F}^{-1}(v))\mathrm{U}(\mathrm{F}^{-1}(w))}{\sigma^2} \\ 0 & v^2 - \frac{\mathrm{U}(\mathrm{F}^{-1}(v))\mathrm{U}(\mathrm{F}^{-1}(w))}{\sigma^2} & vw - \frac{\mathrm{U}(\mathrm{F}^{-1}(w))^2}{\sigma^2} \end{pmatrix}.$$

Beweis Der Beweis verläuft analog zu dem des vorherigen Lemmas. □

In der stochastischen Entwicklung (5.13) kommen nicht nur Z_n, C_{1n} beziehungsweise C_{2n} als zufällige Größen vor, sondern auch X_n, wobei wir bereits wissen, daß

$$X_n \xrightarrow{\mathcal{L}} Y_\infty \qquad \text{für } n \to \infty \tag{5.12}$$

gilt. Obwohl im allgemeinen für $u \in (0,1)$ die Zufallsgrößen $(Z_n, C_{1n}(u), C_{2n}(u))$ und X_n für festes $n \in \mathbb{N}$ nicht unabhängig sind, zeigen wir im Folgenden, daß zumindest asymptotische Unabhängigkeit gegeben ist. Bezeichnet $\lfloor x \rfloor$ für $x \in \mathbb{R}$ den unteren Gaußabschnitt von x, so definieren wir

$$N = n - \lfloor \sqrt{n} \rfloor$$

sowie

$$Z_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{N} X_{i-1} e_i,$$

$$C_{1n}^*(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \big(1_{\{e_i \leq \mathrm{F}^{-1}(u)\}} - u \big) \qquad \text{für } u \in (0,1),$$

$$C_{2n}^*(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \left(1_{\{e_i \leq \mathrm{F}^{-1}(u)\}} - u - \frac{\mathrm{U}(\mathrm{F}^{-1}(u))}{\sigma^2} e_i \right) \qquad \text{für } u \in (0,1)$$

und

$$X_n^* = \sum_{i=N+1}^{n} \rho^{n-i} e_i.$$

Die Zufallsgrößen $(Z_n^*, C_{1n}^*(u), C_{2n}^*(u))$ einerseits und X_n^* andererseits sind offensichtlich für jedes $n \in \mathbb{N}$ unabhängig, und mit diesen Bezeichnungen gilt das folgende Lemma.

Lemma 5.9

Gilt die Bedingung (2.10), so folgt für beliebiges $u \in (0,1)$ die Aussage

$$\begin{pmatrix} Z_n \\ C_{1n}(u) \\ C_{2n}(u) \\ X_n \end{pmatrix} - \begin{pmatrix} Z_n^* \\ C_{1n}^*(u) \\ C_{2n}^*(u) \\ X_n^* \end{pmatrix} \longrightarrow 0 \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.} \tag{5.17}$$

Beweis Wir betrachten zunächst die erste Komponente der Vektoren. Den Bezeichnungen aus Lemma 5.7 folgend gilt

$$Z_n - Z_n^* = \sum_{i=N+1}^{n} Z_{ni},$$

und für beliebiges $\varepsilon > 0$ erhalten wir unter Ausnutzung der Zentriertheit der e_i mit der Tschebyscheff-Ungleichung

$$\mathrm{P}\left(\left| \sum_{i=N+1}^{n} X_{i-1} e_i \right| \geq \sqrt{n}\varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2} \sum_{i=N+1} \mathrm{E}(X_{i-1}^2) \leq \frac{\sigma^4}{(1-\rho^2)\varepsilon^2} \cdot \frac{1}{\sqrt{n}} = o(1),$$

was die Behauptung für die erste Komponente beweist.

Bei der Betrachtung der zweiten Komponente stellen wir fest, daß für beliebiges $u \in (0,1)$ die Verteilungsgleichheit

$$C_{1n}(u) - C_{1n}^*(u) \overset{\mathcal{L}}{=} \frac{\lfloor\sqrt{n}\rfloor^{1/2}}{\sqrt{n}} \cdot \frac{1}{\lfloor\sqrt{n}\rfloor^{1/2}} \sum_{i=1}^{\lfloor\sqrt{n}\rfloor} \left(1_{\{e_i \leq \mathrm{F}^{-1}(u)\}} - u\right)$$

gilt, und die rechte Seite ist offensichtlich von der Ordnung $O_{\mathrm{P}}(n^{-1/4})$ also insbesondere eine stochastische Nullfolge. Gleiches gilt für die dritte Komponente.

Zuletzt stellen wir fest, daß

$$X_n - X_n^* = \sum_{i=1}^{N} \rho^{n-i} e_i = \rho^{(n-N)} X_N$$

gilt, und unter Berücksichtigung von (5.12) konvergiert die rechte Seite stochastisch gegen Null, was den Beweis des Lemmas abschließt. □

Aus den vorangegangenen Lemmata können wir jetzt zusammenfassend die folgende Verteilungskonvergenzaussage folgern.

Lemma 5.10

Es gelte (2.10), und es sei $r \in \{1,2\}$. Dann gilt

$$\begin{pmatrix} Z_n \\ C_{rn}(v) \\ C_{rn}(w) \\ X_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} Z \\ N_{r1} \\ N_{r2} \\ Y_\infty \end{pmatrix} \qquad \text{für } n \to \infty \tag{5.18}$$

mit Y_∞ gemäß (5.12), wobei die Zufallsgrößen Z, (N_{r1}, N_{r2}) sowie Y_∞ unabhängig sind und

$$\begin{pmatrix} Z \\ N_{r1} \\ N_{r2} \end{pmatrix} \overset{\mathcal{L}}{=} \mathcal{N}_3\left(0, \Gamma_r\right)$$

gilt.

Beweis Sei $r \in \{1,2\}$ beliebig. Aus (5.17) und (5.15) beziehungsweise (5.16) folgt

$$\begin{pmatrix} Z_n^* \\ C_{rn}^*(v) \\ C_{rn}^*(w) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} Z \\ N_{r1} \\ N_{r2} \end{pmatrix} \qquad \text{für } n \to \infty,$$

und aus (5.17) und (5.12) erhalten wir

$$X_n^* \xrightarrow{\mathcal{L}} Y_\infty \qquad \text{für } n \to \infty.$$

Nun folgern wir aus der Unabhängigkeit von $(Z_n^*, C_{rn}^*(v), C_{rn}^*(w))$ und X_n^* die Unabhängigkeit von (Z, N_{r1}, N_{r2}) und Y_∞. Die Unabhängigkeit von Z und (N_{r1}, N_{r2}) können wir unmittelbar an der Kovarianzmatrix Γ_r ablesen, da (Z, N_{r1}, N_{r2}) normalverteilt ist. Aus

$$\begin{pmatrix} Z_n \\ C_{rn}(v) \\ C_{rn}(w) \\ X_n \end{pmatrix} = \begin{pmatrix} Z_n^* \\ C_{rn}^*(v) \\ C_{rn}^*(w) \\ X_n^* \end{pmatrix} + o_{\mathrm{P}}(1)$$

gemäß (5.17) folgt die Behauptung mit dem Satz von Cramér-Slutzky. □

Wir haben nun alle Hilfsmittel bereitgestellt, um das Hauptresultat dieses Kapitels zu beweisen.

Satz 5.11

Sei $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ oder $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ und $\hat{I}_n$ gemäß (5.1) definiert. Unter den Voraussetzungen (2.10) und (3.13) gilt

$$\begin{aligned} &\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in \hat{I}_n \middle| \mathcal{A}_n\right) - (1-\alpha)\right) \\ &\xrightarrow{\mathcal{L}} \quad \left(\mathrm{f}\left(\mathrm{F}^{-1}(w)\right) - \mathrm{f}\left(\mathrm{F}^{-1}(v)\right)\right) \frac{1-\rho^2}{\sigma^2} Z Y_\infty \\ &\qquad - \begin{cases} (N_{12} - N_{11}) & \text{falls } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res} \\ (N_{22} - N_{21}) & \text{falls } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z} \end{cases} \qquad \text{für } n \to \infty, \end{aligned} \tag{5.19}$$

wobei die in der Grenzverteilung vorkommenden Zufallsvariablen wie im vorhergehenden Lemma definiert sind.

Beweis Dies folgt aus der stochastischen Entwicklung (5.13) und der Verteilungskonvergenz (5.18) mit Hilfe des Stetigkeitssatzes für Verteilungskonvergenz. □

Wie angekündigt vergleichen wir die Varianzen der beiden (zentrierten) Grenzvariablen, die durch die unterschiedliche Schätzung von F entstehen. Wir bezeichnen die Grenzvariable im Fall $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ mit V_1 und im Fall $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ mit V_2. Durch einfaches Nachrechnen erhält man dann

$$\mathrm{Var}(V_1) = \left(\mathrm{f}\left(\mathrm{F}^{-1}(w)\right) - \mathrm{f}\left(\mathrm{F}^{-1}(v)\right)\right)^2 \sigma^2 + \alpha(1-\alpha)$$

und

$$\mathrm{Var}(V_2) = \left(\mathrm{f}\left(\mathrm{F}^{-1}(w)\right) - \mathrm{f}\left(\mathrm{F}^{-1}(v)\right)\right)^2 \sigma^2 + \alpha(1-\alpha) - \frac{\left(\mathrm{U}\left(\mathrm{F}^{-1}(w)\right) - \mathrm{U}\left(\mathrm{F}^{-1}(v)\right)\right)^2}{\sigma^2},$$

und man sieht sofort, daß stets

$$\mathrm{Var}(V_1) - \mathrm{Var}(V_2) = \frac{\left(\mathrm{U}\left(\mathrm{F}^{-1}(w)\right) - \mathrm{U}\left(\mathrm{F}^{-1}(v)\right)\right)^2}{\sigma^2} \geq 0$$

gilt, so daß die genauere Schätzung von F durch den besseren Schätzer $\mathrm{F}_{n,res,z}$ der Schätzung durch $\mathrm{F}_{n,res}$ vorzuziehen ist.

Allerdings muß beachtet werden, daß im Falle einer symmetrischen Fehlerverteilung F auch die Funktionen U und f symmetrisch sind, so daß in diesem Falle für die Grenzvarianzen

$$\mathrm{Var}(V_1) = \mathrm{Var}(V_2) = \alpha(1-\alpha)$$

gilt. Bei symmetrischer Fehlerverteilung sind also die Schätzer $\mathrm{F}_{n,res}$ und $\mathrm{F}_{n,res,z}$ in diesem Sinne gleichwertig.

Zum Schluß dieses Kapitels betrachten wir noch einseitige Prognoseschranken. Wir werden sehen, daß die mit Hilfe von $\mathrm{F}_{n,res,z}$ konstruierten Schranken auch im Falle einer symmetrischen Fehlerverteilung vorteilhaft gegenüber den auf $\mathrm{F}_{n,res}$ basierenden ist.

Betrachtet man nämlich zu einem vorgegebenen Niveau $\alpha \in (0,1)$ bei beobachteten $X_0, \ldots, X_n$ die (datenabhängige) obere Prognoseschranke

$$\hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(1-\alpha)$$

mit $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ oder $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$, so kann man

$$\mathrm{P}\left(X_{n+1} \leq \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(1-\alpha) | \mathcal{A}_n\right) \longrightarrow 1-\alpha \qquad \text{für } n \to \infty \qquad \text{P-stochastisch}$$

analog zu (5.3) zeigen, und (5.19) entsprechend gilt auch

$$\begin{aligned} &\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \leq \hat{\rho}_n X_n + \hat{\mathrm{F}}_n^{-1}(1-\alpha)|\mathcal{A}_n\right) - (1-\alpha)\right) \\ &\xrightarrow{\mathcal{L}} \mathrm{f}(\mathrm{F}^{-1}(1-\alpha))\frac{1-\rho^2}{\sigma^2} Z Y_\infty - \begin{cases} N_1 & \text{falls } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res} \\ N_2 & \text{falls } \hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z} \end{cases} \qquad \text{für } n \to \infty \end{aligned}$$

mit von (Z, Y_∞) unabhängigen Zufallsvariablen

$$N_1 \stackrel{\mathcal{L}}{=} \mathcal{N}\left(0, \alpha(1-\alpha)\right) \qquad \text{und} \qquad N_2 \stackrel{\mathcal{L}}{=} \mathcal{N}\left(0, \alpha(1-\alpha) - \frac{\mathrm{U}(\mathrm{F}^{-1}(1-\alpha))^2}{\sigma^2}\right).$$

Vergleicht man die Varianzen der Grenzverteilungen, so erhält man offensichtlich als Differenz

$$\mathrm{Var}(N_1) - \mathrm{Var}(N_2) = \frac{\mathrm{U}(\mathrm{F}^{-1}(1-\alpha))^2}{\sigma^2} \geq 0,$$

und diese Differenz ist auch im Fall einer symmetrischen Fehlerverteilung F im allgemeinen ungleich Null, so daß hier in jedem Falle eine Verbesserung zu erwarten ist. Gleiches gilt natürlich für die analoge Konstruktion unterer $1-\alpha$-Prognoseschranken.

Wie sich die unterschiedliche Schätzung von F bei Prognoseintervallen sowie Prognoseschranken in der Praxis auswirkt, untersuchen wir empirisch im letzten Kapitel dieser Arbeit mittels einer Simulationsstudie.

Kapitel 6

Nichtparametrische Zwei-Schritt-Prognoseintervalle in stabilen autoregressiven Prozessen erster Ordnung

6.1 Modellvoraussetzungen und technische Hilfsmittel

Nachdem wir im letzten Kapitel eine Möglichkeit aufgezeigt haben, nichtparametrische Ein-Schritt-Prognoseintervalle in stabilen AR(1)-Prozessen zu konstruieren, schließt sich in natürlicher Weise die Frage nach $1-\alpha$-Prognosen mit einem größeren Schätzhorizont, also nach Intervallen $\hat{I}_{s,n}$ mit

$$\mathrm{P}\left(X_{n+s} \in \hat{I}_{s,n} | \mathcal{A}_n\right) \longrightarrow 1-\alpha \qquad \text{für } n \to \infty \qquad \text{P-stochastisch}$$

und einem $s \geq 2$ an. Wegen

$$X_{n+s} = \rho^s X_n + \sum_{k=1}^{s} \rho^{s-k} e_{n+k}$$

liegt es nahe, zu einem Niveau $\alpha \in (0,1)$ Prognoseintervalle der Form

$$\hat{I}_{s,n} = \left[\hat{\rho}_n^s X_n + \hat{\mathrm{G}}_n^{-1}\left(\frac{\alpha}{2}\right), \hat{\rho}_n^s X_n + \hat{\mathrm{G}}_n^{-1}\left(1-\frac{\alpha}{2}\right)\right]$$

mit einem geeigneten Verteilungsschätzer $\hat{\mathrm{G}}_n$ für die Verteilung von $\sum_{k=1}^{s} \rho^{s-k} e_{n+k}$ zu untersuchen. Dabei benötigt man für alle $s \geq 2$ die gleichen Beweistechniken, weshalb wir uns aus Gründen der Notationsvereinfachung auf $s = 2$ beschränken. Außerdem betrachten wir zunächst nur den Fall $\rho \in (0,1)$, auf die Fälle $\rho \in (-1,0)$ und $\rho = 0$ werden wir am Ende des Kapitels noch eingehen.

Zuerst stellen wir wieder einige technische Hilfsmittel bereit, die wir im Folgenden benötigen werden.

6.1.1 Faltungen

Zu zwei Verteilungsfunktionen H_1 und H_2 ist die *Faltung*

$$(\mathrm{H}_1 * \mathrm{H}_2)(x) = \int_{\mathbb{R}} \mathrm{H}_1(x-y)\mathrm{H}_2(dy) \qquad \text{für } x \in \mathbb{R}$$

definiert, und für zwei unabhängige Zufallsvariablen Y_1 und Y_2, die nach H_1 beziehungsweise H_2 verteilt sind, gilt bekanntermaßen

$$Y_1 + Y_2 \overset{\mathcal{L}}{=} \mathrm{H}_1 * \mathrm{H}_2.$$

Bezeichnen wir mit G die Verteilungsfunktion (beziehungsweise die Verteilung) von $\rho e_1 + e_2$, so gilt

$$\mathrm{G}(x) = \mathrm{P}(\rho e_1 + e_2 \leq x) = (\mathrm{F} * \mathrm{F}(\cdot/\rho))(x) = \mathrm{E}(\mathrm{F}(x - \rho e_1)) = \mathrm{E}\left(\mathrm{F}\left(\frac{x - e_1}{\rho}\right)\right)$$

für beliebiges $x \in \mathbb{R}$.

Für Faltungen gilt das nächste allgemeine Lemma.

Lemma 6.1

Sei $s \in \mathbb{N}$, $\delta > 0$ und seien $\mathrm{H}_1, \ldots, \mathrm{H}_s$ sowie $\mathrm{K}_1, \ldots, \mathrm{K}_s$ Verteilungsfunktionen, dann gelten

$$\sup_{x \in \mathbb{R}} \left|(\mathrm{H}_1 * \cdots * \mathrm{H}_s)(x) - (\mathrm{K}_1 * \cdots * \mathrm{K}_s)(x)\right| \leq \sum_{l=1}^{s} \sup_{x \in \mathbb{R}} |H_l(x) - \mathrm{K}_l(x)| \tag{6.1}$$

und

$$\begin{aligned} \sup_{|r-t| \leq \delta} &|(\mathrm{H}_1 * \cdots * \mathrm{H}_s)(r) - (\mathrm{K}_1 * \cdots * \mathrm{K}_s)(r) \\ &\quad -(\mathrm{H}_1 * \cdots * \mathrm{H}_s)(t) + (\mathrm{K}_1 * \cdots * \mathrm{K}_s)(t)| \\ \leq &\sum_{l=1}^{s} \sup_{|r-t| \leq \delta} |\mathrm{H}_l(r) - \mathrm{K}_l(r) - \mathrm{H}_l(t) + \mathrm{K}_l(r)|. \end{aligned} \tag{6.2}$$

Beweis Der Beweis für beide Aussagen ist elementar, so daß wir nur (6.2) ausführen. Für $s = 1$ ist nichts zu zeigen. Die Behauptung gelte also für ein $s \in \mathbb{N}$. Wir setzen

$$\mathcal{H} = \mathrm{H}_1 * \cdots * \mathrm{H}_s \qquad \text{und} \qquad \mathcal{K} = \mathrm{K}_1 * \cdots * \mathrm{K}_s.$$

Für beliebiges $\delta > 0$ schätzen wir nun wie folgt ab:

$$\begin{aligned}
&\sup_{|r-t|\le\delta} |(\mathcal{H} * \mathrm{H}_{s+1})(r) - (\mathcal{K} * \mathrm{K}_{s+1})(r) - (\mathcal{H} * \mathrm{H}_{s+1})(t) + (\mathcal{K} * \mathrm{K}_{s+1})(t)| \\
&\quad \le \sup_{|r-t|\le\delta} |(\mathcal{H} * \mathrm{H}_{s+1})(r) - (\mathcal{H} * \mathrm{K}_{s+1})(r) - (\mathcal{H} * \mathrm{H}_{s+1})(t) + (\mathcal{H} * \mathrm{K}_{s+1})(t)| \\
&\qquad + \sup_{|r-t|\le\delta} |(\mathcal{H} * \mathrm{K}_{s+1})(r) - (\mathcal{K} * \mathrm{K}_{s+1})(r) - (\mathcal{H} * \mathrm{K}_{s+1})(t) + (\mathcal{K} * \mathrm{K}_{s+1})(t)| \\
&\quad = w_1 + w_2.
\end{aligned}$$

Weiter gilt

$$\begin{aligned}
w_1 &\le \sup_{|r-t|\le\delta} \int_{\mathbb{R}} |\mathrm{H}_{s+1}(r-y) - \mathrm{K}_{s+1}(r-y) - \mathrm{H}_{s+1}(t-y) + \mathrm{K}_{s+1}(t-y)|\, \mathcal{H}(dy) \\
&\le \sup_{|r-t|\le\delta} \sup_{y\in\mathbb{R}} |\mathrm{H}_{s+1}(r-y) - \mathrm{K}_{s+1}(r-y) - \mathrm{H}_{s+1}(t-y) + \mathrm{K}_{s+1}(t-y)| \\
&= \sup_{|r-t|\le\delta} |\mathrm{H}_{s+1}(r) - \mathrm{K}_{s+1}(r) - \mathrm{H}_{s+1}(t) + \mathrm{K}_{s+1}(t)|\,,
\end{aligned}$$

und andererseits impliziert die Induktionsvoraussetzung

$$\begin{aligned}
w_2 &\le \sup_{|r-t|\le\delta} \int_{\mathbb{R}} |\mathcal{H}(r-y) - \mathcal{K}(r-y) - \mathcal{H}(t-y) + \mathcal{K}(t-y)|\, \mathrm{K}_{s+1}(dy) \\
&\le \sup_{|r-t|\le\delta} \sup_{y\in\mathbb{R}} |\mathcal{H}(r-y) - \mathcal{K}(r-y) - \mathcal{H}(t-y) + \mathcal{K}(t-y)| \\
&\le \sum_{l=1}^{s} \sup_{|r-t|\le\delta} |\mathrm{H}_l(r) - \mathrm{K}_l(r) - \mathrm{H}_l(t) + \mathrm{K}_l(r)|,
\end{aligned}$$

was zusammen mit der Abschätzung für w_1 den Beweis von (6.2) abschließt. □

6.1.2 Der zentrale Grenzwertsatz für U-Statistiken und das asymptotische Verhalten von Faltungen

In diesem Abschnitt werden wir einige grundlegende Resultate über U-Statistiken bereitstellen. Sei dazu $\vartheta \in \mathbb{R}$ ein Funktional der Verteilung F und $(e_i)_{i\in\mathbb{N}}$ eine unabhängig und identisch nach F verteilte Folge von Zufallsvariablen. Ist $\mathrm{h} : \mathbb{R}^s \longrightarrow \mathbb{R}$ eine in ihren Argumenten symmetrische Funktion mit $\mathrm{E}(\mathrm{h}(e_1, \ldots, e_s)) = \vartheta$, so heißt h *Kern vom Grad s für* ϑ. Weiter definieren wir die zugehörige U-Statistik durch

$$U_n = \binom{n}{s}^{-1} \sum_{1\le i_1<\cdots<i_s\le n} \mathrm{h}(e_{i_1}, \ldots, e_{i_s}).$$

Die U-Statistik U_n ist ebenfalls ein in den Argumenten symmetrischer und erwartungstreuer Schätzer für ϑ. Den folgenden zentralen Grenzwertsatz für U-Statistiken findet man zum Beispiel bei Lee (1990).

Satz 6.2 *(vgl. Satz 1, Abschnitt 3.2.1 aus Lee (1990))*

Unter der Voraussetzung $\mathrm{E}(\mathrm{h}^2(e_1, \ldots, e_s)) < \infty$ gilt

$$U_n - \vartheta = \frac{s}{n} \sum_{i=1}^{n} \left(\mathrm{h}^{(1)}(e_i) - \vartheta\right) + o_{\mathrm{P}}\left(n^{-1/2}\right) \tag{6.3}$$

mit

$$\vartheta = \mathrm{E}(\mathrm{h}(e_1, \ldots, e_s))$$

und

$$\mathrm{h}^{(1)}(x) = \mathrm{E}(\mathrm{h}(x, e_2, \ldots, e_s)) \qquad \text{für } x \in \mathbb{R}.$$

Insbesondere folgt

$$\sqrt{n}(U_n - \vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, s^2 \mathrm{Var}\left(\mathrm{h}^{(1)}(e_1)\right)\right) \qquad \text{für } n \to \infty.$$

Dabei interpretieren wir Verteilungskonvergenz gegen $\mathcal{N}(0, 0)$ als stochastische Konvergenz gegen Null.

Mit Hilfe dieses Satzes ist es möglich, das asymptotische Verteilungsverhalten von Faltungen empirischer Verteilungsfunktionen zu untersuchen. Zum Beispiel gilt

$$\begin{aligned}(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho)(x) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} 1_{\{\rho e_i + e_j \leq x\}} = O_{\mathrm{P}}\left(n^{-1}\right) \\ &\quad + \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \frac{1}{2} \left(1_{\{\rho e_{i_1} + e_{i_2} \leq x\}} + 1_{\{\rho e_{i_2} + e_{i_1} \leq x\}}\right)\end{aligned} \tag{6.4}$$

für $x \in \mathbb{R}$, und die letzte Summe in obiger Gleichungskette ist für jedes $x \in \mathbb{R}$ gerade eine U-Statistik mit dem Kern

$$\mathrm{h}_x(y, z) = \frac{1}{2} \left(1_{\{\rho y + z \leq x\}} + 1_{\{\rho z + y \leq x\}}\right) \qquad \text{für } y, z \in \mathbb{R} \tag{6.5}$$

und

$$\mathrm{E}(\mathrm{h}_x(e_1, e_2)) = (\mathrm{F} * \mathrm{F}(\cdot/\rho))(x) = \mathrm{G}(x),$$

so daß wir für jedes $x \in \mathbb{R}$ die Verteilungskonvergenz

$$\sqrt{n}\left[(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho)(x) - \mathrm{G}(x)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\ 4 \cdot \mathrm{Var}\left(\mathrm{h}_x^{(1)}(e_1)\right)\right) \tag{6.6}$$

mit

$$\mathrm{h}_x^{(1)}(t) = \frac{1}{2}\left(\mathrm{F}(x-\rho t) + \mathrm{F}\left(\frac{x-t}{\rho}\right)\right) \qquad \text{für } t \in \mathbb{R} \tag{6.7}$$

folgern können.

In einigen später betrachteten Anwendungen der Theorie der U-Statistiken hängt der Kern der U-Statistik zusätzlich auch noch von weiteren Parametern ab. In dieser Situation benötigen wir eine gleichmäßige Version von (6.3), es soll also für eine ganze Klasse $\mathcal{H}$ von Kernen vom Grad s beziehungsweise die zugehörigen U-Statistiken

$$\sup_{\mathrm{h}\in\mathcal{H}} \left| U_n(\mathrm{h}) - \mathrm{E}(\mathrm{h}(e_1,\ldots,e_s)) - \frac{s}{n}\sum_{i=1}^{n}\left(\mathrm{h}^{(1)}(e_i) - \mathrm{E}(\mathrm{h}(e_1,\ldots,e_s))\right)\right| = o_{\mathrm{P}}\left(n^{-1/2}\right)$$

gelten. Geeignete Funktionenklassen $\mathcal{H}$ sind zum Beispiel *Vapnik-Červonenkis-Subgraph-Klassen* von Funktionen. Die folgenden Definitionen und Vorgehensweisen sind Kapitel 5 in de la Peña und Giné (1999) entnommen.

Eine Klasse $\mathcal{C}$ von Untermengen einer Menge M heißt *Vapnik-Červonenkis-Klasse,* falls eine Zahl $k \in \mathbb{N}$ existiert, so daß für alle Teilmengen $A \subseteq M$ mit $|A| = k$ die echte Ungleichung

$$|\{A \cap C : C \in \mathcal{C}\}| < 2^k$$

gilt. Anders ausgedrückt ist die Klasse $\mathcal{C}$ nicht in der Lage, sämtliche Teilmengen irgend einer k-elementigen Teilmenge A von M „auszuschneiden." Die kleinste Zahl k, für die obige Ungleichung gilt, heißt *VC-Index* von $\mathcal{C}$.

Für VC-Klassen $\mathcal{C}$ und $\mathcal{D}$ sind auch die Klassen

$$\begin{aligned} \mathcal{C}\cup\mathcal{D} &= \{C \cup D : C \in \mathcal{C},\ D \in \mathcal{D}\}, \\ \mathcal{C}\cap\mathcal{D} &= \{C \cap D : C \in \mathcal{C},\ D \in \mathcal{D}\} \\ \text{und}\quad \mathcal{C}\times\mathcal{D} &= \{C \times D : C \in \mathcal{C},\ D \in \mathcal{D}\} \end{aligned}$$

VC-Klassen. Außerdem ist eine beliebige Teilklasse $\mathcal{C}$ einer VC-Klasse $\mathcal{D}$ stets wieder eine VC-Klasse.

Ein einfaches Beispiel einer VC-Klasse ist die Klasse aller abgeschlossenen Halbebenen in $\mathbb{R}^2$. Mit elementargeometrischen Argumenten rechnet man leicht nach, daß sie den VC-Index Vier besitzt. Auch die Klasse der offenen Halbebenen in $\mathbb{R}^2$ ist dann natürlich eine VC-Klasse mit VC-Index Vier.

Eine Klasse $\mathcal{F}$ von reellwertigen Funktionen auf M heißt *Vapnik-Červonenkis-Subgraph-Klasse* (kurz VC-SG-Klasse), falls die Klasse

$$\mathcal{C} = \{C_{\mathrm{f}} : \mathrm{f} \in \mathcal{F}\}$$

der Subgraphen

$$C_{\mathrm{f}} = \{(m,t) \in M \times \mathbb{R} : 0 \le t \le \mathrm{f}(m) \quad \text{oder} \quad \mathrm{f}(m) \le t \le 0\}$$

eine VC-Klasse ist.

Weiter heißt eine Klasse $\mathcal{F}$ von reellwertigen, auf dem meßbaren Raum $(M, \mathcal{M})$ definierten Funktionen *meßbar*, falls es einen vollständig metrisierbaren, separablen Raum S mit zugehöriger Borel-σ-Algebra $\mathcal{B}(S)$ und eine surjektive Abbildung $\mathrm{T} : S \longrightarrow \mathcal{F}$ gibt, so daß die Abbildung $\tau : S \times M \longrightarrow \mathbb{R}$ mit

$$\tau(s, m) = \mathrm{T}(s)(m) \qquad \text{für } s \in S, m \in M$$

$\mathcal{B}(S) \otimes \mathcal{M} - \mathcal{B}^*$-meßbar ist.

Der folgende Satz verallgemeinert den obigen zentralen Grenzwertsatz für U-Statistiken auf VC-SG-Klassen von Kernen.

Satz 6.3 *(vgl. Satz 5.3.3 von de la Peña und Giné (1999))*

Sei $\mathcal{H}$ eine meßbare VC-SG-Klasse symmetrischer Kerne vom Grad s. Ist

$$\mathrm{E}\left(\sup_{h \in \mathcal{H}} \left|\mathrm{h}^2(e_1, \ldots, e_s)\right|\right) < \infty,$$

so gilt

$$\sup_{\mathrm{h} \in \mathcal{H}} \Bigg| U_n(\mathrm{h}) - \mathrm{E}(\mathrm{h}(e_1, \ldots, e_s)) \\ - \frac{s}{n} \sum_{i=1}^{n} \left(\mathrm{h}^{(1)}(e_i) - \mathrm{E}(\mathrm{h}(e_1, \ldots, e_s))\right) \Bigg| = o_{\mathrm{P}}\left(n^{-1/2}\right). \tag{6.8}$$

Beweis Dies ergibt sich aus Satz 5.3.3 und dem Beweis zu Satz 5.3.1 in de la Peña und Giné (1999). □

6.2 Das asymptotische Verhalten nichtparametrischer Prognoseintervalle

Wegen (5.10) ist der Kleinst-Quadrate-Schätzer (KQS) $\hat{\rho}_n$ insbesondere ein konsistenter Schätzer für ρ, so daß

$$\mathrm{P}\left(\{\hat{\rho}_n \leq 0\} \cup \{\hat{\rho}_n \geq 1\}\right) \leq 2\mathrm{P}\left(|\hat{\rho}_n - \rho| \geq \min\left\{\frac{\rho}{2}, \frac{1-\rho}{2}\right\}\right) \longrightarrow 0 \qquad \text{für } n \to \infty$$

gilt, weshalb wir im Folgenden stets $\hat{\rho}_n \in (0,1)$ annehmen. Zur Konstruktion nichtparametrischer Zwei-Schritt-Prognoseintervalle zu einem Niveau $\alpha \in (0,1)$ definieren wir G wie im Abschnitt 6.1.1 und

$$\mathrm{G}_{n,res} = \mathrm{F}_{n,res} * \mathrm{F}_{n,res}(\cdot/\hat{\rho}_n) \qquad \text{sowie} \qquad \mathrm{G}_{n,res,z} = \mathrm{F}_{n,res,z} * \mathrm{F}_{n,res,z}(\cdot/\hat{\rho}_n).$$

Weiter bezeichnen G^{-1}, $\mathrm{G}_{n,res}^{-1}$ und $\mathrm{G}_{n,res,z}^{-1}$ wieder die zugehörigen Quantilfunktionen sowie

$$a = \mathrm{G}^{-1}(v) \qquad \text{und} \qquad b = \mathrm{G}^{-1}(w)$$

mit v, w wie in Kapitel 5. Schließlich definieren wir zu $\hat{\mathrm{G}}_n = \mathrm{G}_{n,res}$ oder $\hat{\mathrm{G}}_n = \mathrm{G}_{n,res,z}$ das nichtparametrische $1-\alpha$-Zwei-Schritt-Prognoseintervall

$$\hat{I}_n = \left[\ \hat{\rho}_n^2 X_n + \hat{\mathrm{G}}_n^{-1}(v), \hat{\rho}_n^2 X_n + \hat{\mathrm{G}}_n^{-1}(w)\ \right]. \tag{6.9}$$

Wir werden nun wie in Kapitel 5 zeigen, daß die so konstruierten Zwei-Schritt-Prognoseintervalle asymptotisch das Niveau einhalten, wobei die auf $\mathrm{F}_{n,res,z}$ basierenden wieder in einem gewissen Sinne genauer sind als die aus $\mathrm{F}_{n,res}$ hergeleiteten.

6.2.1 Konsistenz

Bei stetiger Verteilungsfunktion F ist auch G stetig, so daß

$$\begin{aligned} \mathrm{P}\left(X_{n+2} \in \hat{I}_n | \mathcal{A}_n\right) &= \mathrm{G}\left((\hat{\rho}_n^2 - \rho^2)X_n + \hat{\mathrm{G}}_n^{-1}(w)\right) \\ &\quad -\mathrm{G}\left((\hat{\rho}_n^2 - \rho^2)X_n + \hat{\mathrm{G}}_n^{-1}(v)\right) \end{aligned} \tag{6.10}$$

gilt. Zum Beweis der Konsistenz zeigen wir zunächst das folgende Lemma.

Lemma 6.4

Unter den Voraussetzungen (2.10) und (3.13) gilt

$$(\hat{\rho}_n^2 - \rho^2)X_n + \hat{\mathrm{G}}_n^{-1}(u) \longrightarrow \mathrm{G}^{-1}(u) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch} \tag{6.11}$$

für beliebiges $u \in (0,1)$.

Beweis Sei $u \in (0,1)$. Für den ersten Summanden gilt

$$(\hat{\rho}_n^2 - \rho^2)X_n = (\hat{\rho}_n - \rho)(\hat{\rho}_n + \rho)X_n = O_{\mathrm{P}}(n^{-1/2})$$

gemäß (5.10) und (5.12). Es muß also nur noch

$$\hat{\mathrm{G}}_n^{-1}(u) \longrightarrow \mathrm{G}^{-1}(u) \qquad \text{für } n \to \infty \qquad \text{P-stochastisch} \tag{6.12}$$

gezeigt werden. Auf Grund von (3.13) besitzt mit F auch G eine $\lambda - fs$ positive Dichte g, und G^{-1} ist somit auf ganz $(0,1)$ stetig. Lemma 5.1 impliziert also (6.12), falls wir

$$\sup_{x \in \mathbb{R}} |\hat{\mathrm{G}}_n(x) - \mathrm{G}(x)| = o_{\mathrm{P}}(1)$$

zeigen können.

Wegen (6.1) gilt

$$\sup_{x\in\mathbb{R}} |\hat{\mathrm{G}}_n(x) - \mathrm{G}(x)| \leq \sup_{x\in\mathbb{R}} |\hat{\mathrm{F}}_n(x) - \mathrm{F}(x)| + \sup_{x\in\mathbb{R}} |\hat{\mathrm{F}}_n(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)|,$$

und der erste Summand konvergiert gemäß (3.14) beziehungsweise (3.18) und Glivenko-Cantelli stochastisch gegen Null.

Im Falle $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res}$ gilt für den zweiten Summanden die Abschätzung

$$\begin{aligned}\sup_{x\in\mathbb{R}} |\mathrm{F}_{n,res}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| \quad &\leq \quad \sup_{x\in\mathbb{R}} |\mathrm{F}_{n,res}(x/\hat{\rho}_n) - \mathrm{F}_n(x/\hat{\rho}_n)| \\ &\quad + \sup_{x\in\mathbb{R}} |\mathrm{F}_n(x/\hat{\rho}_n) - \mathrm{F}(x/\hat{\rho}_n)| \\ &\quad + \sup_{x\in\mathbb{R}} |\mathrm{F}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| \\ &= \quad S_1 + S_2 + S_3,\end{aligned}$$

und S_1 und S_2 konvergieren auf Grund von (3.14) und Glivenko-Cantelli stochastisch gegen Null, weshalb wir nur noch S_3 untersuchen müssen.

Sei dazu $\varepsilon > 0$ beliebig und $B > 0$ so groß, daß

$$\mathrm{F}(-B) \leq \varepsilon \qquad \text{und} \qquad \mathrm{F}(B) \geq 1 - \varepsilon$$

gilt. Wegen $\hat{\rho}_n < 1$ gelten für alle $x < -B$ die Abschätzungen

$$\mathrm{F}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho) \leq \mathrm{F}(x/\hat{\rho}_n) \leq \mathrm{F}(-B) \leq \varepsilon$$

und

$$\mathrm{F}(x/\rho) - \mathrm{F}(x/\hat{\rho}_n) \leq \mathrm{F}(x/\rho) \leq \mathrm{F}(-B) \leq \varepsilon,$$

also

$$\sup_{x<-B} |\mathrm{F}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| \leq \varepsilon,$$

und ebenso folgert man

$$\sup_{x>B} |\mathrm{F}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| \leq \varepsilon.$$

Andererseits ist f als gleichmäßig stetige Dichte beschränkt, so daß wegen der Konsistenz des KQS

$$\sup_{|x|\leq B} |\mathrm{F}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| \leq B \left|\frac{1}{\hat{\rho}_n} - \frac{1}{\rho}\right| \sup_{x\in\mathbb{R}} |\mathrm{f}(x)| = o_{\mathrm{P}}(1)$$

und insgesamt

$$S_3 \leq o_{\mathrm{P}}(1) + \varepsilon$$

gilt.

Ist $\hat{\mathrm{F}}_n = \mathrm{F}_{n,res,z}$ so gilt

$$\begin{aligned}\sup_{x\in\mathbb{R}} |\mathrm{F}_{n,res,z}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| &\le \sup_{x\in\mathbb{R}} |\mathrm{F}_{n,res,z}(x/\hat{\rho}_n) - \mathrm{F}(x/\hat{\rho}_n)| \\ &\quad + \sup_{x\in\mathbb{R}} |\mathrm{F}(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| \\ &= R_1 + R_2.\end{aligned}$$

Der Term R_2 ist S_3 von oben, und R_1 konvergiert wegen (3.18) und Glivenko-Cantelli stochastisch gegen Null.

Es gilt also in beiden Fällen

$$\sup_{x\in\mathbb{R}} |\hat{\mathrm{F}}_n(x/\hat{\rho}_n) - \mathrm{F}(x/\rho)| = o_{\mathrm{P}}(1),$$

und Lemma 5.1 impliziert die Behauptung. □

Nun können wir die Konsistenz der Intervalle $\hat{I}_n$ folgern.

Satz 6.5

Unter den Voraussetzungen (2.10) und (3.13) gilt

$$\mathrm{P}\left(X_{n+2} \in \hat{I}_n | \mathcal{A}_n\right) \longrightarrow 1-\alpha \qquad \text{für } n\to\infty \qquad \text{P-stochastisch.} \tag{6.13}$$

Beweis Dies folgt aus dem vorangegangen Lemma, der Stetigkeit von G und der Darstellung (6.10). □

Damit ist bewiesen, daß die mittels $\hat{\mathrm{F}}_n$ beziehungsweise $\hat{\mathrm{G}}_n$ konstruierten Zwei-Schritt-Prognoseintervalle konsistent sind. Wegen (6.11) konvergieren ihre Endpunkte stochastisch gegen den gleichen Grenzwert, so daß ihre Länge asymptotisch gleich ist.

6.2.2 Verteilungskonvergenz bedingter Überdeckungswahrscheinlichkeiten

In diesem Abschnitt werden wir analog zu Kapitel 5 die Zufallsvariablen

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+2} \in \hat{I}_n | \mathcal{A}_n\right) - (1-\alpha)\right)$$

untersuchen und insbesondere die Unterschiede aufzeigen, die aus der unterschiedlichen Schätzung von G durch $\mathrm{G}_{n,res}$ beziehungsweise $\mathrm{G}_{n,res,z}$ folgen. Auch hierzu benötigen wir zunächst einige vorbereitende Lemmata. Zur Erinnerung wiederholen wir an dieser Stelle auch nochmals die Definition

$$V_\delta(x) = \{y \in \mathbb{R} : |x-y| \le \delta\} \qquad \text{für } x \in \mathbb{R}, \delta > 0.$$

Lemma 6.6

Es gelte (2.10) und (3.13). Für beliebiges $x \in \mathbb{R}$ und $C > 0$ gilt

$$\begin{aligned} & \left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/(\rho + un^{-1/2})\right)\right)(y) - \left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/\rho\right)\right)(y) \\ = \;& -un^{-1/2}\mathrm{E}\left(\mathrm{f}(y-\rho e_1)e_1\right) + R_n(u,y) \end{aligned} \tag{6.14}$$

mit

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n}\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}} |R_n(u,y)| \ge \varepsilon\right) = 0.$$

Beweis Setzen wir

$$\mathrm{h}_{u,y}(s,t) = \frac{1}{2}\left(1_{\left\{(\rho+un^{-1/2})s+t\le y\right\}} + 1_{\left\{(\rho+un^{-1/2})t+s\le y\right\}}\right) \qquad \text{für } s,t,y\in\mathbb{R},\ |u|\le C,$$

so gilt

$$\mathrm{E}\left(\mathrm{h}_{u,y}(e_1,e_2)\right) = \mathrm{E}\left(\mathrm{F}\left(y-(\rho+un^{-1/2})e_1\right)\right) = \mathrm{E}\left(\mathrm{F}\left(\frac{y-e_1}{\rho+un^{-1/2}}\right)\right).$$

Weiter definieren wir

$$\mathrm{E}_1(y) = \mathrm{E}(\mathrm{f}(y-\rho e_1)e_1) \qquad \text{für } y\in\mathbb{R} \tag{6.15}$$

und

$$\mathrm{E}_2(y) = \mathrm{E}\left(\mathrm{f}\left(\frac{y-e_1}{\rho}\right)(y-e_1)\right) = \rho^2\mathrm{E}_1(y) \qquad \text{für } y\in\mathbb{R}. \tag{6.16}$$

Es ist

$$U_n(\mathrm{h}_{u,y}) = \binom{n}{2}^{-1}\sum_{1\le i<j\le n}\mathrm{h}_{u,y}(e_i,e_j)$$

die zu $\mathrm{h}_{u,y}$ gehörige U-Statistik, und für beliebiges $\delta > 0$ gilt

$$\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}}\left|\left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/(\rho+un^{-1/2})\right)\right)(y) - U_n(\mathrm{h}_{u,y})\right| \le \frac{2}{n},$$

was man analog zu (6.4) zeigt. Für jedes $\varepsilon > 0$ genügt es demnach

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathrm{P}\left(\sqrt{n}\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}}\left|U_n(\mathrm{h}_{u,y}) - U_n(\mathrm{h}_{0,y}) + \frac{u}{\sqrt{n}}\mathrm{E}_1(y)\right| \ge \varepsilon\right) = 0 \tag{6.17}$$

zu beweisen. Für beliebiges $|u| \leq C$ und $y \in \mathrm{I\!R}$ gilt offensichtlich

$$\begin{aligned} U_n(\mathrm{h}_{u,y}) &= \left\{ \frac{2}{n} \sum_{i=1}^{n} \mathrm{h}_{u,y}^{(1)}(e_i) - \mathrm{E}(\mathrm{h}_{u,y}(e_1, e_2)) \right\} \\ &\quad + \left\{ U_n(\mathrm{h}_{u,y}) - \mathrm{E}(\mathrm{h}_{u,y}(e_1, e_2)) - \frac{2}{n} \sum_{i=1}^{n} \left(\mathrm{h}_{u,y}^{(1)}(e_i) - \mathrm{E}(\mathrm{h}_{u,y}(e_1, e_2)) \right) \right\} \\ &= \frac{2}{n} \sum_{i=1}^{n} \mathrm{h}_{u,y}^{(1)}(e_i) - \mathrm{E}(\mathrm{h}_{u,y}(e_1, e_2)) + R_{1n}(u, y). \end{aligned}$$

Wir wählen ein $\varepsilon > 0$ mit $\{\rho - \varepsilon, \rho + \varepsilon\} \subset (0, 1)$, und definieren durch

$$\mathrm{h}'_{r,y}(s, t) = \frac{1}{2} \left(1_{\{(\rho+r)s+t \leq y\}} + 1_{\{(\rho+r)t+s \leq y\}} \right) \qquad \text{für } s, t, y \in \mathrm{I\!R},\ |r| \leq \varepsilon$$

eine Klasse von symmetrischen Kernen vom Grad Zwei. Für beliebiges $\delta > 0$ und alle großen n gilt

$$\begin{aligned} &\sup_{\substack{|u| \leq C \\ y \in V_\delta(x)}} |R_{1n}(u, y)| \\ &\leq \sup_{\substack{|r| \leq \varepsilon \\ y \in \mathrm{I\!R}}} \left| U_n(\mathrm{h}'_{r,y}) - \mathrm{E}(\mathrm{h}'_{r,y}(e_1, e_2)) - \frac{2}{n} \sum_{i=1}^{n} \left({\mathrm{h}'_{r,y}}^{(1)}(e_i) - \mathrm{E}(\mathrm{h}'_{r,y}(e_1, e_2)) \right) \right|, \end{aligned}$$

und hieraus erhalten wir mittels (6.8), daß

$$\sup_{\substack{|u| \leq C \\ y \in V_\delta(x)}} |R_{1n}(u, y)| = o_{\mathrm{P}}(n^{-1/2})$$

gilt. Offensichtlich ist nämlich

$$\mathrm{E} \left(\sup_{\substack{|r| \leq \varepsilon \\ y \in \mathrm{I\!R}}} \left| \mathrm{h}'^{2}_{r,y}(e_1, \ldots, e_s) \right| \right) \leq 1,$$

deshalb genügt es für (6.8) zu zeigen, daß

$$\mathcal{H} = \{\mathrm{h}'_{r,y} : |r| \leq \varepsilon, y \in \mathrm{I\!R}\}$$

eine meßbare VC-SG-Klasse ist.

Wir zeigen zunächst die Meßbarkeit. Die Menge $[-\varepsilon, \varepsilon] \times \mathrm{I\!R}$ ist offensichtlich ein vollständiger, separabler metrischer Raum und die Abbildung $\mathrm{T} : [-\varepsilon, \varepsilon] \times \mathrm{I\!R} \longrightarrow \mathcal{H}$ mit

$$\mathrm{T}(r, y) = \mathrm{h}'_{r,y} \qquad \text{für } (r, y) \in [-\varepsilon, \varepsilon] \times \mathrm{I\!R}$$

ist offensichtlich surjektiv. Schließlich ist $\tau : [-\varepsilon, \varepsilon] \times \mathrm{I\!R}^3 \longrightarrow \mathrm{I\!R}$ mit

$$\tau(r, y, s_1, s_2) = \mathrm{h}'_{r,y}(s_1, s_2) \qquad \text{für } (r, y, s_1, s_2) \in [-\varepsilon, \varepsilon] \times \mathrm{I\!R}^3$$

als Komposition stetiger Funktionen mit der Indikatorfunktion eines abgeschlossenen Halbstrahls offensichtlich $\mathcal{B}([-\varepsilon, \varepsilon]) \otimes \mathcal{B}_3^* - \mathcal{B}^*$-meßbar.

Es bleibt zu zeigen, daß $\mathcal{H}$ eine VC-SG-Klasse ist. Sei dazu $(r, y) \in [-\varepsilon, \varepsilon] \times \mathrm{I\!R}$. Dann ist

$$\begin{aligned} C_{\mathrm{h}'_{r,y}} &= \{(s_1, s_2) \in \mathrm{I\!R}^2 : \mathrm{h}'_{r,y}(s_1, s_2) = 1\} \times [0, 1] \\ &\quad \cup \{(s_1, s_2) \in \mathrm{I\!R}^2 : \mathrm{h}'_{r,y}(s_1, s_2) = 1/2\} \times [0, 1/2] \\ &\quad \cup \{(s_1, s_2) \in \mathrm{I\!R}^2 : \mathrm{h}'_{r,y}(s_1, s_2) = 0\} \times \{0\} \end{aligned}$$

und

$$\mathcal{C} = \{C_{\mathrm{h}'_{r,y}} : |r| \le \varepsilon, y \in \mathrm{I\!R}\}.$$

Wir erinnern nochmals daran, daß Schnitte, Vereinigungen, kartesische Produkte und Teilklassen von VC-Klassen wieder VC-Klassen sind. Die einelementigen Klassen $\{[0, 1]\}$, $\{[0, 1/2]\}$ und $\{\{0\}\}$ sind offensichtlich VC-Klassen mit VC-Index Eins, so daß es genügt zu zeigen, daß die Klassen

$$\{\{(s_1, s_2) \in \mathrm{I\!R}^2 : h'_{r,y}(s_1, s_2) = p\} : |r| \le \varepsilon, y \in \mathrm{I\!R}\} \qquad \text{für } p = 0, 1/2, 1$$

VC-Klassen sind. Dies wiederum gilt, denn die Klassen

$$\begin{aligned} &\{\{(s_1, s_2) \in \mathrm{I\!R}^2 : (\rho + r)s_1 + s_2 \le y\} : |r| \le \varepsilon, y \in \mathrm{I\!R}\}, \\ &\{\{(s_1, s_2) \in \mathrm{I\!R}^2 : s_1 + (\rho + r)s_2 \le y\} : |r| \le \varepsilon, y \in \mathrm{I\!R}\}, \\ &\{\{(s_1, s_2) \in \mathrm{I\!R}^2 : (\rho + r)s_1 + s_2 > y\} : |r| \le \varepsilon, y \in \mathrm{I\!R}\} \\ \text{und}\quad &\{\{(s_1, s_2) \in \mathrm{I\!R}^2 : s_1 + (\rho + r)s_2 > y\} : |r| \le \varepsilon, y \in \mathrm{I\!R}\} \end{aligned}$$

sind jeweils VC-Klassen, da sie sämtlich Teilklassen aller abgeschlossenen oder aller offenen Halbebenen in $\mathrm{I\!R}^2$ sind. Insgesamt ist $\mathcal{C}$ eine VC-Klasse und somit gilt

$$\sup_{\substack{|u| \le C \\ y \in V_\delta(x)}} |R_{1n}(u, y)| = o_{\mathrm{P}}(n^{-1/2}).$$

Auf Grund der Definition von $\mathrm{h}_{u,y}$ gilt weiter

$$\begin{aligned} \frac{2}{n} \sum_{i=1}^n \left(\mathrm{h}^{(1)}_{u,y}(e_i) - \mathrm{h}^{(1)}_{0,y}(e_i)\right) &= \frac{1}{n} \sum_{i=1}^n \left(\mathrm{F}(y - (\rho + un^{-1/2})e_i) - \mathrm{F}(y - \rho e_i)\right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(\mathrm{F}\left(\frac{y - e_i}{\rho + un^{-1/2}}\right) - \mathrm{F}\left(\frac{y - e_i}{\rho}\right)\right) \\ &= S_{1n}(u, y) + S_{2n}(u, y), \end{aligned}$$

und der Mittelwertsatz liefert

$$S_{1n}(u,y) = -\frac{u}{\sqrt{n}} \cdot \frac{1}{n}\sum_{i=1}^{n} \mathrm{f}(y-\rho e_i)e_i + \frac{1}{n}\sum_{i=1}^{n}\int_{y-\rho e_i}^{y-(\rho+un^{-1/2})e_i} \mathrm{f}(s) - \mathrm{f}(y-\rho e_i)ds.$$

Wegen der gleichmäßigen Stetigkeit von f existiert zu jedem $\varepsilon > 0$ ein $\delta' > 0$, so daß

$$|\mathrm{f}(s) - \mathrm{f}(t)| \leq \frac{\varepsilon}{2C\mathrm{E}(|e_1|)} \qquad \text{für } |s-t| \leq \delta'$$

gilt. Außerdem gilt für alle großen n die Ungleichung

$$\sup_{|u|\leq C} \left|\frac{un^{-1/2}}{\rho(\rho+un^{-1/2})}\right| \leq \frac{2C}{\rho^2\sqrt{n}} = O_{\mathrm{P}}(n^{-1/2}) \tag{6.18}$$

$$\text{sowie} \qquad \sup_{|u|\leq C} \left|\frac{u}{\sqrt{n}}\right| \leq \frac{C}{\sqrt{n}} = O_{\mathrm{P}}(n^{-1/2}), \tag{6.19}$$

und auf dem Ereignis $\{\frac{1}{n}\sum_{i=1}^{n}|e_i| \leq 2\mathrm{E}(|e_1|)\} \cap \{\max_{1\leq i\leq n}|e_i| \leq \frac{\delta' n^{1/2}}{C}\}$ folgt für alle n und für alle $\delta > 0$ die Abschätzung

$$\sqrt{n} \sup_{\substack{|u|\leq C \\ y\in V_\delta(x)}} \left|\frac{1}{n}\sum_{i=1}^{n}\int_{y-\rho e_i}^{y-(\rho+un^{-1/2})e_i} \mathrm{f}(s) - \mathrm{f}(y-\rho e_i)ds\right| \leq \varepsilon.$$

Das Gegenereignis aber hat auf Grund des starken Gesetzes der großen Zahlen und wegen $\max_{1\leq i\leq n}|e_i| = o_{\mathrm{P}}(n^{1/2})$ gegen Null konvergierende Wahrscheinlichkeit.

Ebenso impliziert der Mittelwertsatz, daß

$$\begin{aligned} S_{2n}(u,y) &= -\frac{un^{-1/2}}{\rho(\rho+un^{-1/2})} \cdot \frac{1}{n}\sum_{i=1}^{n} \mathrm{f}\left(\frac{y-e_i}{\rho}\right)(y-e_i) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\int_{\frac{y-e_i}{\rho}}^{\frac{y-e_i}{\rho+un^{-1/2}}} \mathrm{f}(s) - \mathrm{f}\left(\frac{y-e_i}{\rho}\right)ds \end{aligned}$$

gilt. Sei nun $0 < \delta \leq 1$ und $\varepsilon > 0$ beliebig. Dann existiert ein $\delta' > 0$, so daß

$$|\mathrm{f}(s) - \mathrm{f}(t)| \leq \frac{\varepsilon\rho^2}{2C(|x|+1+2\mathrm{E}(|e_1|))} \qquad \text{für } |s-t| \leq \delta'$$

gilt, da f gleichmäßig stetig ist.

Auf dem Ereignis $\{\frac{1}{n}\sum_{i=1}^{n}|e_i| \leq 2\mathrm{E}(|e_1|)\} \cap \{|x| + 1 + \max_{1\leq i\leq n}|e_i| \leq \frac{\delta'\rho^2 n^{1/2}}{2C}\}$ können wir für alle großen n wie oben

$$\sqrt{n} \sup_{\substack{|u|\leq C \\ y\in V_\delta(x)}} \left|\frac{1}{n}\sum_{i=1}^{n}\int_{\frac{y-e_i}{\rho}}^{\frac{y-e_i}{\rho+un^{-1/2}}} \mathrm{f}(s) - \mathrm{f}\left(\frac{y-e_i}{\rho}\right)ds\right| \leq \varepsilon$$

folgern, während das Gegenereignis wegen $\max_{1\le i\le n}|e_i| = o_{\mathrm{P}}(n^{1/2})$ und dem starken Gesetz der großen Zahlen wieder gegen Null konvergierende Wahrscheinlichkeit besitzt.

Außerdem gilt

$$\begin{aligned} & \mathrm{E}(\mathrm{h}_{u,y}(e_1,e_2)) - \mathrm{E}(\mathrm{h}_{0,y}(e_1,e_2)) \\ = \; & -\frac{un^{-1/2}}{\rho(\rho+un^{-1/2})}\mathrm{E}_2(y) + R_{2n}(u,y) \end{aligned}$$

mit

$$\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}} |R_{2n}(u,y)| = o(n^{-1/2})$$

für alle $0<\delta\le 1$, was man mit ähnlichen Argumenten zeigt.

Als nächstes beweisen wir, daß

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathrm{P}\left(\sup_{y\in V_\delta(x)}\left|\frac{1}{n}\sum_{i=1}^n \mathrm{f}\left(\frac{y-e_i}{\rho}\right)(y-e_i) - \mathrm{E}_2(y)\right| \ge \varepsilon\right) = 0 \qquad (6.20)$$

gilt. Dazu schätzen wir zunächst

$$\begin{aligned} & \sup_{y\in V_\delta(x)}\left|\frac{1}{n}\sum_{i=1}^n \mathrm{f}\left(\frac{y-e_i}{\rho}\right)(y-e_i) - \mathrm{E}_2(y)\right| \\ \le \; & \sup_{y\in V_\delta(x)}\left|\frac{1}{n}\sum_{i=1}^n \left(\mathrm{f}\left(\frac{y-e_i}{\rho}\right)(y-e_i) - \mathrm{f}\left(\frac{x-e_i}{\rho}\right)(x-e_i)\right)\right| \\ & + \left|\frac{1}{n}\sum_{i=1}^n \mathrm{f}\left(\frac{x-e_i}{\rho}\right)(x-e_i) - \mathrm{E}_2(x)\right| \\ & + \sup_{y\in V_\delta(x)} |\mathrm{E}_2(x) - \mathrm{E}_2(y)| \\ = \; & A_{1n}(\delta) + A_{2n} + A_3(\delta) \end{aligned}$$

ab, und zeigen die Behauptung für jeden Summanden einzeln.

Sei also $\varepsilon > 0$. Dann gibt es wegen der gleichmäßigen Stetigkeit von f ein $\delta' > 0$, so daß

$$|\mathrm{f}(s) - \mathrm{f}(t)| \le \frac{\varepsilon}{8\mathrm{E}(|x-e_1|)} \qquad \text{für } |s-t| \le \delta'$$

gilt. Ist weiter $0 < \delta \le \min\{\rho\delta', \frac{\varepsilon}{4\sup_{t\in\mathbb{R}}|\mathrm{f}(t)|}\}$, so gilt für ein solches δ auf dem Ereignis $\{\frac{1}{n}\sum_{i=1}^n |x-e_i| \le 2\mathrm{E}(|x-e_1|)\}$ die Abschätzung

$$A_{1n}(\delta) \;\le\; \sup_{y\in V_\delta(x)} \frac{1}{n}\sum_{i=1}^n \left|\mathrm{f}\left(\frac{y-e_i}{\rho}\right) - \mathrm{f}\left(\frac{x-e_i}{\rho}\right)\right| |x-e_i|$$

$$
\begin{aligned}
& + \sup_{y \in V_\delta(x)} \frac{1}{n} \sum_{i=1}^{n} \left| \mathrm{f}\left(\frac{y-e_i}{\rho}\right) \right| |y-x| \\
\leq \; & \frac{\varepsilon}{8\mathrm{E}(|x-e_1|)} \frac{1}{n} \sum_{i=1}^{n} |x-e_i| + \sup_{t \in \mathbb{R}} |\mathrm{f}(t)| \delta \\
\leq \; & \frac{\varepsilon}{2},
\end{aligned}
$$

und somit können wir für solche δ auch

$$
\limsup_{n \to \infty} \mathrm{P}(A_{1n}(\delta) \geq \varepsilon) \leq \limsup_{n \to \infty} \mathrm{P}\left(\frac{1}{n} \sum_{i=1}^{n} |x-e_i| > 2\mathrm{E}(|x-e_1|)\right) = 0
$$

aus dem starken Gesetz der großen Zahlen folgern.

Das starke Gesetz der großen Zahlen impliziert weiter

$$
A_{2n} = o_{\mathrm{P}}(1),
$$

so daß nur noch $A_3(\delta)$ zu untersuchen bleibt.

Für dieselben $\delta > 0$ wie bei der Betrachtung von $A_{1n}(\delta)$ gilt aber auch

$$
\mathrm{A}_3(\delta) \leq \frac{\varepsilon}{2},
$$

was man mit den gleichen Argumenten wie dort nachrechnet. Somit ist (6.20) gezeigt.

Analog kann man

$$
\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P}\left(\sup_{y \in V_\delta(x)} \left| \frac{1}{n} \sum_{i=1}^{n} \mathrm{f}(y - \rho e_i) - \mathrm{E}_1(y) \right| \geq \varepsilon\right) = 0 \tag{6.21}
$$

für beliebiges $\varepsilon > 0$ zeigen, so daß wir aus (6.18), (6.19), (6.20) und (6.21) insgesamt

$$
S_{1n}(u,y) + S_{2n}(u,y) - (\mathrm{E}(\mathrm{h}_{u,y}(e_1,e_2)) - \mathrm{E}(\mathrm{h}_{0,y}(e_1,e_2))) = -\frac{u}{\sqrt{n}} \mathrm{E}_1(y) + R_{3n}(u,y)
$$

mit

$$
\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P}\left(\sqrt{n} \sup_{\substack{|u| \leq C \\ y \in V_\delta(x)}} |R_{3n}(u,y)| \geq \varepsilon\right) = 0
$$

für alle $\varepsilon > 0$ folgern können.

Zusammenfügen der bewiesenen Aussagen ergibt

$$
\begin{aligned}
& \sup_{\substack{|u| \leq C \\ y \in V_\delta(x)}} \left| U_n(\mathrm{h}_{u,y}) - U_n(\mathrm{h}_{0,y}) + \frac{u}{\sqrt{n}} \mathrm{E}_1(y) \right| \\
= \; & \sup_{\substack{|u| \leq C \\ y \in V_\delta(x)}} |R_{1n}(u,y) - R_{1n}(0,y) + R_{3n}(u,y)|,
\end{aligned}
$$

so daß (6.17) folgt, was noch zu zeigen war. □

Auch das nächste Lemma ist von ähnlichem Typ. Es stellt die Verbindung zwischen Faltungen basierend auf F_n und solchen basierend auf $\mathrm{F}_{n,z}$ her.

Lemma 6.7

Unter den Voraussetzungen (2.10) und (3.13) gilt für jedes $x \in \mathbb{R}$ und jedes $C > 0$ die Aussage

$$\begin{aligned}&\left(\mathrm{F}_{n,z} * \mathrm{F}_{n,z}(\cdot/\rho + un^{-1/2})\right)(y) - \left(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho + un^{-1/2})\right)(y)\\ &= \; -\left[\mathrm{E}\left(\mathrm{U}\left(\frac{y-e_1}{\rho}\right)\right) + \mathrm{E}(\mathrm{U}(y-\rho e_1))\right]\frac{1}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^n e_i + R_n(u,y) \qquad (6.22)\end{aligned}$$

mit

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathrm{P}\left(\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}} |R_n(u,y)| \ge \varepsilon\right) = 0$$

für beliebiges $\varepsilon > 0$.

Beweis Wir definieren zunächst

$$\mathrm{U}_1(y) = \mathrm{E}(\mathrm{U}(y-\rho e_1)) = \mathrm{E}\left(e_1 \mathrm{F}\left(\frac{y-e_1}{\rho}\right)\right) \qquad \text{für } y \in \mathbb{R} \qquad (6.23)$$

und

$$\mathrm{U}_2(y) = \mathrm{E}\left(\mathrm{U}\left(\frac{y-e_1}{\rho}\right)\right) = \mathrm{E}\left(e_1 \mathrm{F}\left(y-\rho e_1\right)\right) \qquad \text{für } y \in \mathbb{R}. \qquad (6.24)$$

Nun verwenden wir die Definition der Faltung und die stochastische Entwicklung (2.16), um

$$\begin{aligned}&\left(\mathrm{F}_{n,z} * \mathrm{F}_{n,z}(\cdot/\rho + un^{-1/2})\right)(y)\\ &= \; \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} 1_{\left\{e_i \le \frac{y-s}{\rho+un^{-1/2}}\right\}} - \mathrm{U}\left(\frac{y-s}{\rho+un^{-1/2}}\right)\frac{e_i}{\sigma^2}\mathrm{F}_{n,z}(ds)\\ &\quad + \int_{\mathbb{R}} R_{1n}\left(\frac{y-s}{\rho+un^{-1/2}}\right)\mathrm{F}_{n,z}(ds)\end{aligned}$$

mit

$$\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}} \left|R_{1n}\left(\frac{y-s}{\rho+un^{-1/2}}\right)\right| \le \sup_{t\in\mathbb{R}} |R_{1n}(t)| = o_{\mathrm{P}}\left(n^{-1/2}\right)$$

zu erhalten. Da $\mathrm{F}_{n,z}$ eine Verteilungsfunktion ist, hat das zweite Integral damit für beliebiges $\delta > 0$ die Ordnung $o_\mathrm{p}(n^{-1/2})$. Ausintegrieren des ersten Integrals bezüglich $\mathrm{F}_{n,z}$ ergibt

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_{\mathrm{I\!R}} 1_{\left\{e_i \le \frac{y-s}{\rho+un^{-1/2}}\right\}} - \mathrm{U}\left(\frac{y-s}{\rho+un^{-1/2}}\right)\frac{e_i}{\sigma^2}\mathrm{F}_{n,z}(ds)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{1+t_n e_j}1_{\{(\rho+un^{-1/2})e_i+e_j\le y\}}\\
&\quad -\left(\frac{1}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n}e_i\right)\frac{1}{n}\sum_{j=1}^{n}\frac{1}{1+t_n e_j}\mathrm{U}\left(\frac{y-e_j}{\rho+un^{-1/2}}\right)\\
&= A_{1n}(u,y) - A_{2n}(u,y),
\end{aligned}$$

und wegen

$$\frac{1}{1+s} = 1 - s + \frac{s^2}{1+s} \qquad \text{für } s \neq -1$$

können wir

$$\begin{aligned}
&A_{1n}(u,y) - \left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/\rho + un^{-1/2}\right)\right)(y)\\
&= -t_n\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}e_j 1_{\{(\rho+un^{-1/2})e_i+e_j\le y\}}\\
&\quad +t_n^2\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{e_j^2}{1+t_n e_j}1_{\{(\rho+un^{-1/2})e_i+e_j\le y\}}\\
&= -S_{1n}(u,y) + R_{2n}(u,y)
\end{aligned}$$

folgern. Dabei gilt für den Restterm R_{2n} unter Beachtung von

$$\max_{1\le i\le n}\left|\frac{1}{1+t_n e_i} - 1\right| = o_\mathrm{P}(1), \tag{6.25}$$

was in Lemma 2.2 von Genz (2001) bewiesen wird, offensichtlich

$$\sup_{\substack{|u|\le C\\ y\in V_\delta(x)}} |R_{2n}(u,y)| \le t_n^2 \max_{1\le i\le n}\frac{1}{1+t_n e_i}\cdot\frac{1}{n}\sum_{i=1}^{n}e_i^2 = O_\mathrm{P}\left(n^{-1}\right).$$

Andererseits erhalten wir

$$\begin{aligned}
S_{1n}(u,y) &= t_n\mathrm{U}_1(y)\\
&\quad +t_n\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}e_j 1_{\{e_j\le y-(\rho-un^{-1/2})e_i\}} - \mathrm{U}(y-(\rho+un^{-1/2})e_i)\right)
\end{aligned}$$

$$\begin{aligned}
&+t_n \frac{1}{n} \sum_{i=1}^{n} (\mathrm{U}(y - (\rho + un^{-1/2})e_i) - \mathrm{U}(y - \rho e_i)) \\
&+t_n \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{U}(y - \rho e_i) - \mathrm{U}_1(y) \right) \\
= \quad & t_n \mathrm{U}_1(y) + t_n (R_{3n}(u,y) + R_{4n}(u,y) + R_{5n}(u,y)),
\end{aligned}$$

und wegen $t_n = O_{\mathrm{P}}(n^{-1/2})$ zeigen wir nun

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P} \left(\sup_{\substack{|u| \le C \\ y \in V_\delta(x)}} |R_{rn}(u,y)| \ge \varepsilon \right) = 0$$

für $r \in \{3,4,5\}$ und beliebiges $\varepsilon > 0$.

Wegen Lemma 2.2 in Zhang (1997) gilt

$$\sup_{\substack{|u| \le C \\ y \in V_\delta(x)}} |R_{3n}(u,y)| \le \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^{n} e_j 1_{\{e_j \le t\}} - \mathrm{U}(t) \right| = o_{\mathrm{P}}(1)$$

für jedes $\delta > 0$.

Weiter gibt es auf Grund der gleichmäßigen Stetigkeit von U zu jedem $\varepsilon > 0$ ein $\delta' > 0$, so daß

$$|\mathrm{U}(s) - \mathrm{U}(t)| \le \frac{\varepsilon}{2} \qquad \text{für } |s - t| \le \delta'$$

gilt. Auf dem Ereignis $\{\max_{1 \le i \le n} |e_i| \le \frac{\delta' n^{1/2}}{C}\}$ erhalten wir folglich

$$\sup_{\substack{|u| \le C \\ y \in V_\delta(x)}} |R_{4n}(u,y)| \le \frac{\varepsilon}{2},$$

und somit gilt

$$\mathrm{P} \left(\sup_{\substack{|u| \le C \\ y \in V_\delta(x)}} |R_{4n}(u,y)| \ge \varepsilon \right) \le \mathrm{P} \left(\max_{1 \le i \le n} |e_i| > \frac{\delta' n^{1/2}}{C} \right) = o(1)$$

für beliebiges $\delta > 0$ wegen $\max_{1 \le i \le n} |e_i| = o_{\mathrm{P}}(n^{1/2})$.

Der Restterm $R_{5n}(u,y) = R_{5n}(y)$ schließlich ist unabhängig von u, und es gilt

$$\begin{aligned}
&\sup_{y \in V_\delta(x)} \left| \frac{1}{n} \sum_{i=1}^{n} \mathrm{U}(y - \rho e_i) - \mathrm{U}_1(y) \right| \\
\le \quad & \sup_{y \in V_\delta(x)} \frac{1}{n} \sum_{i=1}^{n} |\mathrm{U}(y - \rho e_i) - \mathrm{U}(x - \rho e_i)|
\end{aligned}$$

$$\begin{aligned}
&+\left|\frac{1}{n}\sum_{i=1}^{n} \mathrm{U}(x-\rho e_i) - \mathrm{U}_1(x)\right| \\
&+\sup_{y\in V_\delta(x)} \mathrm{E}(|\mathrm{U}(x-\rho e_1) - \mathrm{U}(y-\rho e_1)|) \\
=\;& B_{1n}(\delta) + B_{2n} + B_3(\delta).
\end{aligned}$$

Für beliebiges $\varepsilon > 0$ existiert ein $\delta_0 > 0$, so daß

$$|\mathrm{U}(s) - \mathrm{U}(t)| \leq \frac{\varepsilon}{3} \qquad \text{für } |s-t| \leq \delta_0$$

gilt, da U gleichmäßig stetig ist. Somit gilt für alle $0 < \delta \leq \delta_0$ und alle n auch

$$B_{1n}(\delta) \leq \frac{\varepsilon}{3}$$

und

$$B_3(\delta) \leq \frac{\varepsilon}{3}.$$

Der Term B_{2n} aber konvergiert gemäß dem starken Gesetz der großen Zahlen für jedes $\delta > 0$ stochastisch gegen Null, so daß wir für $0 < \delta \leq \delta_0$ schließlich

$$\limsup_{n\to\infty} \mathrm{P}\left(\sup_{y\in V_\delta(x)} |R_{5n}(y)| \geq \varepsilon\right) \leq \limsup_{n\to\infty} \mathrm{P}\left(B_{2n} \geq \frac{\varepsilon}{3}\right) = 0$$

folgern können.

Auf Grund der Beschränktheit von U und (2.13) können wir weiter

$$S_{1n}(u,y) = \mathrm{U}_1(y)\frac{1}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n} e_i + R_{6n}(u,y)$$

ableiten, wobei der Restterm die gewünschte Ordnung hat.

Analog zeigt man

$$A_{2n}(u,y) = \left(\frac{1}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n} e_i\right)\mathrm{U}_2(y) + R_{7n}(u,y)$$

mit

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathrm{P}\left(\sup_{\substack{|u|\leq C\\ y\in V_\delta(x)}} |R_{7n}(u,y)| \geq \varepsilon\right) = 0$$

für jedes $\varepsilon > 0$.

Insgesamt ergibt sich

$$\begin{aligned}
&A_{1n}(u,y) - A_{2n}(u,y) - \left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/\rho + un^{-1/2}\right)\right)(y) \\
= \;& -\left(\frac{1}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^n e_i\right)(\mathrm{U}_1(y)+\mathrm{U}_2(y)) \\
&+R_{2n}(u,y) - R_{6n}(u,y) - R_{7n}(u,y),
\end{aligned}$$

und das Zusammenfassen der erhaltenen Ergebnisse schließt den Beweis des Lemmas ab. □

Als nächstes geben wir stochastische Entwicklungen für $\mathrm{G}_{n,res}$ und $\mathrm{G}_{n,res,z}$ an.

Lemma 6.8

Es gelte (2.10) und (3.13). Für beliebiges $x \in \mathbb{R}$ gilt dann

$$\begin{aligned}
\sqrt{n}\,(\mathrm{G}_{n,res}(x) - \mathrm{G}(x)) \;=\; & \frac{2}{\sqrt{n}}\sum_{i=1}^n \left(\mathrm{h}_x^{(1)}(e_i) - \mathrm{G}(x)\right) \\
& -\sqrt{n}\,(\hat\rho_n - \rho)\,\mathrm{E}_1(x) + o_{\mathrm{P}}(1) \qquad (6.26)
\end{aligned}$$

mit E_1 und $\mathrm{h}_x^{(1)}$ gemäß den Defintionen (6.15) und (6.7).

Beweis Es ist

$$\begin{aligned}
&\sqrt{n}\,(\mathrm{G}_{n,res}(x) - \mathrm{G}(x)) = \sqrt{n}\,(\mathrm{G}_{n,res}(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat\rho_n))(x)) \\
&\quad +\sqrt{n}\,((\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat\rho_n))(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) + (\hat\rho_n - \rho)\mathrm{E}_1(x)) \\
&\quad -\sqrt{n}\,(\hat\rho_n - \rho)\,\mathrm{E}_1(x) + \sqrt{n}\,(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) - \mathrm{G}(x))\,,
\end{aligned}$$

und aus (6.1) und (3.14) folgt

$$|\mathrm{G}_{n,res}(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat\rho_n))(x)| = o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Andererseits gilt für jedes $C > 0$ auf dem Ereignis $\{n^{1/2}|\hat\rho_n - \rho| \le C\}$ die Abschätzung

$$\begin{aligned}
&|(\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat\rho_n))(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) + (\hat\rho_n - \rho)\,\mathrm{E}_1(x)| \\
&\quad \le \sup_{|u|\le C}\left|\left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/(\rho + un^{-1/2})\right)\right)(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) + un^{-1/2}\mathrm{E}_1(x)\right|,
\end{aligned}$$

und die rechte Seite der Ungleichung ist gemäß (6.14) von der Ordnung $o_{\mathrm{P}}(n^{-1/2})$, während

$$\limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n}|\hat\rho_n - \rho| > C\right) \longrightarrow 0 \qquad \text{für } C \to \infty$$

gilt. Schließlich ist

$$(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) = U_n(\mathrm{h}_x) + o_{\mathrm{P}}(n^{-1/2})$$

mit h_x gemäß (6.5), und der zentrale Grenzwertsatz für U-Statistiken schließt den Beweis des Lemmas ab. □

Ein analoges Lemma zeigen wir für den Schätzer $\mathrm{G}_{n,res,z}$.

Lemma 6.9

Es gelte (2.10) und (3.13). Für beliebiges $x \in \mathbb{R}$ gilt dann

$$\begin{aligned}\sqrt{n}\,(\mathrm{G}_{n,res,z}(x) - \mathrm{G}(x)) &= \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathrm{h}_x^{(1)}(e_i) - \mathrm{G}(x)\right)\\ &-(\mathrm{U}_1(x) + \mathrm{U}_2(x))\frac{1}{\sigma^2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i - \sqrt{n}\,(\hat\rho_n - \rho)\,\mathrm{E}_1(x) + o_{\mathrm{P}}(1) \qquad (6.27)\end{aligned}$$

mit $\mathrm{U}_1, \mathrm{U}_2$ gemäß den Definitionen (6.23) und (6.24).

Beweis Es ist

$$\begin{aligned}\sqrt{n}\,(\mathrm{G}_{n,res,z}(x) - \mathrm{G}(x)) &= \sqrt{n}\,(\mathrm{G}_{n,res,z}(x) - (\mathrm{F}_{n,z} * \mathrm{F}_{n,z}(\cdot/\hat\rho_n))(x))\\ &+\sqrt{n}\Bigg((\mathrm{F}_{n,z} * \mathrm{F}_{n,z}(\cdot/\hat\rho_n))(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat\rho_n))(x)\\ &+(\mathrm{U}_1(x) + \mathrm{U}_2(x))\frac{1}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n} e_i\Bigg)\\ &+\sqrt{n}\,((\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat\rho_n))(x) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) + (\hat\rho_n - \rho)\mathrm{E}_1(x))\\ &-(\mathrm{U}_1(x) + \mathrm{U}_2(x))\frac{1}{\sigma^2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i - \sqrt{n}\,(\hat\rho_n - \rho)\,\mathrm{E}_1(x)\\ &+\sqrt{n}\,(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) - \mathrm{G}(x))\,,\end{aligned}$$

und die Diskussion der Restterme erfolgt wie im letzten Lemma mit (2.14) und (3.18) anstelle von (3.14) und unter zusätzlicher Beachtung von (6.22). □

Bevor wir uns wieder unserem eigentlichen Problem, dem Verteilungsverhalten der Zufallsfolgen

$$\sqrt{n}(\mathrm{P}(X_{n+2} \in \hat I_n|\mathcal{A}_n) - (1-\alpha))$$

mit gemäß (6.9) definierten Intervallen $\hat I_n$ zuwenden können, benötigen wir noch zwei Lemmata, die uns lokale Straffheitseigenschaften der stochastischen Prozesse $(n^{1/2}(\mathrm{G}_{n,res} - \mathrm{G}))_{n\in\mathbb{N}}$ und $(n^{1/2}(\mathrm{G}_{n,res,z} - \mathrm{G}))_{n\in\mathbb{N}}$ sicherstellen.

Lemma 6.10

Sei $x \in \mathbb{R}$ und $\varepsilon > 0$. Unter den Voraussetzungen (2.10) und (3.13) gilt dann

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n} \sup_{y \in V_\delta(x)} |\mathrm{G}_{n,res}(y) - \mathrm{G}(y) - (\mathrm{G}_{n,res}(x) - \mathrm{G}(x))| \geq \varepsilon\right) = 0. \quad (6.28)$$

Beweis Wegen (5.11) gibt es ein $C > 0$, so daß

$$\limsup_{n\to\infty} \mathrm{P}(\sqrt{n}|\hat{\rho}_n - \rho| > C) \leq \frac{\varepsilon}{3} \quad (6.29)$$

gilt. Auf dem Ereignis $\{n^{1/2}|\hat{\rho}_n - \rho| \leq C\}$ aber gilt für jedes $\delta > 0$ die Abschätzung

$$\begin{aligned}
&\sqrt{n} \sup_{y \in V_\delta(x)} |\mathrm{G}_{n,res}(y) - \mathrm{G}(y) - (\mathrm{G}_{n,res}(x) - \mathrm{G}(x))| \\
\leq\ & 2\sqrt{n} \sup_{y \in \mathbb{R}} |\mathrm{G}_{n,res}(y) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\hat{\rho}_n))(y)| \\
& + 2\sqrt{n} \sup_{\substack{|u| \leq C \\ y \in V_\delta(x)}} \left| \left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/\rho + un^{-1/2}\right)\right)(y) - (\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(y) + \frac{u}{\sqrt{n}} \mathrm{E}_1(y) \right| \\
& + C \sup_{y \in V_\delta(x)} |\mathrm{E}_1(y) - \mathrm{E}_1(x)| \\
& + \sqrt{n} \sup_{|x-y| \leq \delta} |(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(y) - \mathrm{G}(y) - [(\mathrm{F}_n * \mathrm{F}_n(\cdot/\rho))(x) - \mathrm{G}(x)]| \\
=\ & R_{1n} + R_{2n}(\delta) + R_3(\delta) + R_{4n}(\delta),
\end{aligned}$$

und wir betrachten die Restterme einzeln.

Gemäß (6.1) gilt für beliebiges $\delta > 0$ offensichtlich

$$R_{1n} \leq 4\sqrt{n} \sup_{y \in \mathbb{R}} |\mathrm{F}_{n,res}(y) - \mathrm{F}_n(y)|$$

und wegen (3.14) somit

$$\limsup_{n\to\infty} \mathrm{P}\left(R_{1n} \geq \frac{\varepsilon}{4}\right) = 0. \quad (6.30)$$

Wegen (6.14) gibt es ein $\delta_0 > 0$ so daß

$$\limsup_{n\to\infty} \mathrm{P}\left(R_{2n}(\delta) \geq \frac{\varepsilon}{4}\right) \leq \frac{\varepsilon}{3} \quad (6.31)$$

für alle $0 < \delta \leq \delta_0$ gilt.

Weiter gibt es auf Grund der gleichmäßigen Stetigkeit von f ein $0 < \delta_1 \leq \delta_0$ mit

$$|\mathrm{f}(s) - \mathrm{f}(t)| \leq \frac{\varepsilon}{4C\mathrm{E}(|e_1|)} \qquad \text{für } |s - t| \leq \delta_1,$$

und mit (6.15) gilt folglich

$$R_3(\delta) \leq \frac{\varepsilon}{4} \tag{6.32}$$

für $0 < \delta \leq \delta_1$.

Schließlich gilt wegen (6.2) noch

$$R_{4n}(\delta) \leq 2\sqrt{n} \sup_{|x-y|\leq \delta/\rho} |\mathrm{F}_n(y) - \mathrm{F}(y) - (\mathrm{F}_n(x) - \mathrm{F}(x))|,$$

und die bekannte Straffheit des empirischen Prozesses liefert die Existenz einer Zahl $0 < \delta_2 \leq \delta_1$ mit

$$\limsup_{n\to\infty} \mathrm{P}\left(R_{4n}(\delta) \geq \frac{\varepsilon}{4}\right) \leq \frac{\varepsilon}{3} \tag{6.33}$$

für beliebiges $0 < \delta \leq \delta_2$.

Insgesamt können wir also für alle $0 < \delta \leq \delta_2$ folgern, daß

$$\begin{aligned}
&\limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n} \sup_{y\in V_\delta(x)} |\mathrm{G}_{n,res}(y) - \mathrm{G}(y) - (\mathrm{G}_{n,res}(x) - \mathrm{G}(x))| \geq \varepsilon\right) \\
\leq\;& \limsup_{n\to\infty} \mathrm{P}\left(R_{1n} + R_{2n}(\delta) + R_{4n}(\delta) \geq \frac{3\varepsilon}{4}\right) + \limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n}|\hat{\rho}_n - \rho| > C\right) \\
\leq\;& \varepsilon
\end{aligned}$$

gemäß (6.30), (6.31), (6.32) und (6.33) und (6.29) gilt, was den Beweis des Lemmas abschließt. □

Lemma 6.11

Sei $x \in \mathbb{R}$ und $\varepsilon > 0$. Unter den Voraussetzungen (2.10) und (3.13) gilt dann

$$\begin{aligned}
\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathrm{P}\Bigg(\sqrt{n} \sup_{y\in V_\delta(x)} &|\mathrm{G}_{n,res,z}(y) - \mathrm{G}(y) \\
&-(\mathrm{G}_{n,res,z}(x) - \mathrm{G}(x))| \geq \varepsilon\Bigg) = 0.
\end{aligned} \tag{6.34}$$

Beweis Der Beweis verläuft analog zu dem des vorherigen Lemmas mit (2.14) und (3.18) anstelle von (3.14). Außerdem sind zusätzlich die Terme

$$\begin{aligned}
2\sqrt{n} \sup_{\substack{|u|\leq C \\ y\in V_\delta(x)}} \Bigg| &\left(\mathrm{F}_{n,z} * \mathrm{F}_{n,z}\left(\cdot/\rho + un^{-1/2}\right)\right)(y) \\
&- \left(\mathrm{F}_n * \mathrm{F}_n\left(\cdot/\rho + un^{-1/2}\right)\right)(y) + \left(\frac{1}{\sigma^2} \cdot \frac{1}{n}\sum_{i=1}^n e_i\right)(\mathrm{U}_1(y) + \mathrm{U}_2(y))\Bigg|
\end{aligned}$$

und

$$\sup_{y\in V_\delta(x)} \left| \frac{1}{\sigma^2} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i \right| |\mathrm{U}_2(y) + \mathrm{U}_1(y) - \mathrm{U}_2(x) - \mathrm{U}_1(x)|$$

zu betrachten. Dabei hat der erste wegen (6.22) und der zweite wegen

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i = O_{\mathrm{P}}(1)$$

und der gleichmäßigen Stetigkeit von U die gewünschte Straffheitseigenschaft. □

Wir sind nun in der Lage, eine Entwicklung für die Zufallsvariablen

$$\sqrt{n}(\mathrm{P}(X_{n+2} \in \hat{I}_n | \mathcal{A}_n) - (1-\alpha))$$

anzugeben, die der in (5.13) entspricht.

Lemma 6.12

Es gelte (2.10) und (3.13). Für

$$a = \mathrm{G}^{-1}(v) = \mathrm{G}^{-1}\left(\frac{\alpha}{2}\right), \qquad b = \mathrm{G}^{-1}(w) = \mathrm{G}^{-1}\left(1 - \frac{\alpha}{2}\right)$$

und

$$I_{n,res} = \left[\ \hat{\rho}_n^2 X_n + \mathrm{G}_{n,res}^{-1}(v), \hat{\rho}_n^2 X_n + \mathrm{G}_{n,res}^{-1}(w)\ \right]$$

gilt

$$\begin{aligned} &\sqrt{n}\left(\mathrm{P}\left(X_{n+2} \in I_{n,res} | \mathcal{A}_n\right) - (1-\alpha)\right) \\ &= \ \left\{(\mathrm{g}(b) - \mathrm{g}(a))\, 2\rho X_n + (\mathrm{E}_1(b) - \mathrm{E}_1(a))\right\} \frac{1-\rho^2}{\sigma^2} Z_n \\ &\quad - \left(\frac{2}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathrm{h}_b^{(1)}(e_i) - w\right) - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathrm{h}_a^{(1)}(e_i) - v\right) \right) + o_{\mathrm{P}}(1), \end{aligned} \tag{6.35}$$

wobei g die Dichte von G und Z_n gemäß (5.14), sowie $\mathrm{E}_1, \mathrm{E}_2$ gemäß (6.15) und (6.16) definiert ist.

Beweis Wie beim Beweis zu (5.13) stellen wir zunächst fest, daß wegen der Stetigkeit von G stets

$$\mathrm{P}\left(X_{n+2} \in I_{n,res} | \mathcal{A}_n\right) = \mathrm{G}\left(\left(\hat{\rho}_n^2 - \rho^2\right) X_n + \mathrm{G}_{n,res}^{-1}(w)\right) - \mathrm{G}\left(\left(\hat{\rho}_n^2 - \rho^2\right) X_n + \mathrm{G}_{n,res}^{-1}(v)\right)$$

und

$$1 - \alpha = w - v = \mathrm{G}(b) - \mathrm{G}(a)$$

gilt. Hieraus folgt für $(u,x) \in \{(v,a),(w,b)\}$ mit der Differenzierbarkeit von G weiter

$$\mathrm{G}\left(\left(\hat{\rho}_n^2-\rho^2\right)X_n+\mathrm{G}_{n,res}^{-1}(u)\right)-u=\left(\left(\hat{\rho}_n^2-\rho^2\right)X_n+\mathrm{G}_{n,res}^{-1}(u)-x\right)(\mathrm{g}(x)+o_{\mathrm{P}}(1)).$$

Wegen (6.26) gilt insbesondere

$$\sqrt{n}\left(\mathrm{G}_{n,res}(x)-\mathrm{G}(x)\right)=O_{\mathrm{P}}(1).$$

Dies zusammen mit (6.28) erlaubt uns, den Satz von Bahadur anzuwenden, und wir erhalten wie in Kapitel 5

$$\mathrm{G}_{n,res}^{-1}(u)-x=-\frac{\mathrm{G}_{n,res}(x)-\mathrm{G}(x)}{\mathrm{g}(x)}+o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Das widerum impliziert

$$\begin{aligned}&\mathrm{G}\left(\left(\hat{\rho}_n^2-\rho^2\right)X_n+\mathrm{G}_{n,res}^{-1}(u)\right)-u\\ =\;&\mathrm{g}(x)\left(\hat{\rho}_n^2-\rho^2\right)X_n-\left(\mathrm{G}_{n,res}(x)-\mathrm{G}(x)\right)+o_{\mathrm{P}}\left(n^{-1/2}\right),\end{aligned}\tag{6.36}$$

und aus der Konsistenz des KQS und (5.8) folgt

$$\sqrt{n}\mathrm{g}(x)\left(\hat{\rho}_n^2-\rho^2\right)X_n=\mathrm{g}(x)2\rho\frac{1-\rho^2}{\sigma^2}Z_nX_n+o_{\mathrm{P}}(1).$$

Die Entwicklung (6.26) ergibt mit (5.8) insbesondere

$$\sqrt{n}(\mathrm{G}_{n,res}(x)-\mathrm{G}(x))=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathrm{h}_x^{(1)}(e_i)-\mathrm{G}(x)\right)-\frac{1-\rho^2}{\sigma^2}Z_n\mathrm{E}_1(x)+o_{\mathrm{P}}(1),$$

so daß man die Behauptung des Lemmas erhält, wenn man die beiden zuletzt bewiesenen Aussagen in (6.36) einsetzt. □

Genauso erhalten wir eine entsprechende Entwicklung im Falle der Verwendung des EL-Schätzers $\mathrm{G}_{n,res,z}$.

Lemma 6.13

Es gelte (2.10) und (3.13). Für

$$I_{n,res,z}=\left[\;\hat{\rho}_n^2X_n+\mathrm{G}_{n,res,z}^{-1}(v),\hat{\rho}_n^2X_n+\mathrm{G}_{n,res,z}^{-1}(w)\;\right]$$

gilt

$$\begin{aligned}&\sqrt{n}\left(\mathrm{P}\left(X_{n+2}\in I_{n,res,z}|\mathcal{A}_n\right)-(1-\alpha)\right)\\ =\;&\left\{\left(\mathrm{g}(b)-\mathrm{g}(a)\right)2\rho X_n+\left(\mathrm{E}_1(b)-\mathrm{E}_1(a)\right)\right\}\frac{1-\rho^2}{\sigma^2}Z_n\\ &-\left(\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathrm{h}_b^{(1)}(e_i)-w\right)-\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathrm{h}_a^{(1)}(e_i)-v\right)\right)\\ &+\left\{\mathrm{U}_1(b)+\mathrm{U}_2(b)-\left(\mathrm{U}_1(a)+\mathrm{U}_2(a)\right)\right\}\frac{1}{\sigma^2}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i+o_{\mathrm{P}}(1),\end{aligned}\tag{6.37}$$

mit $\mathrm{U}_1,\mathrm{U}_2$ gemäß (6.23) und (6.24).

Beweis Der Beweis verläuft analog zu dem des letzten Lemmas mit (6.27) und (6.34) anstelle von (6.26) und (6.28). □

Das nächste Lemma stellt die gemeinsame Verteilungskonvergenz aller für uns relevanten Zufallsfolgen sicher.

Lemma 6.14

Sei

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} e_i,$$

$$C_n(x) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathrm{h}_x^{(1)}(e_i) - \mathrm{G}(x)\right) \qquad \text{für } x \in \mathbb{R}$$

und Z_n gemäß (5.14). Unter der Voraussetzung (2.10) gilt dann

$$\begin{pmatrix} Z_n \\ A_n \\ C_n(b) \\ C_n(a) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{pmatrix} \stackrel{\mathcal{L}}{=} \mathcal{N}_4\,(0, \Gamma) \qquad \text{für } n \to \infty,$$

und

$$\Gamma = \begin{pmatrix} \frac{\sigma^4}{1-\rho^2} & 0 & 0 & 0 \\ 0 & \sigma^2 & \mathrm{U}_1(b) + \mathrm{U}_2(b) & \mathrm{U}_1(a) + \mathrm{U}_2(a) \\ 0 & \mathrm{U}_1(b) + \mathrm{U}_2(b) & 4\mathrm{Var}\left(\mathrm{h}_b^{(1)}(e_1)\right) & 4\mathrm{Cov}\left(\mathrm{h}_a^{(1)}(e_1), \mathrm{h}_b^{(1)}(e_1)\right) \\ 0 & \mathrm{U}_1(a) + \mathrm{U}_2(a) & 4\mathrm{Cov}\left(\mathrm{h}_a^{(1)}(e_1), \mathrm{h}_b^{(1)}(e_1)\right) & 4\mathrm{Var}\left(\mathrm{h}_a^{(1)}(e_1)\right) \end{pmatrix}$$

ist die Varianzmatrix der Grenzverteilung. Insbesondere sind Z_n einerseits und $(A_n, C_n(b), C_n(a))$ andererseits asymptotisch unabhängig.

Beweis Die Aussage folgt aus dem mehrdimensionalen zentralen Grenzwertsatz für Martingaldifferenzschemata angewandt auf das vierdimensionale Martingaldifferenzschema $(X_{ni}, \mathcal{A}_i)_{1 \le i \le n}$ gegeben durch

$$X_{ni} = \begin{pmatrix} Z_{ni} \\ A_{ni} \\ C_{ni}(b) \\ C_{ni}(a) \end{pmatrix} = \begin{pmatrix} \frac{X_{i-1}e_i}{\sqrt{n}} \\ \frac{e_i}{\sqrt{n}} \\ \frac{2}{\sqrt{n}}\left(\mathrm{h}_b^{(1)}(e_i) - w\right) \\ \frac{2}{\sqrt{n}}\left(\mathrm{h}_a^{(1)}(e_i) - v\right) \end{pmatrix} \qquad \text{für } 1 \le i \le n.$$

Dazu sind die konditionierte Normierungsbedingung und die komponentenweise konditionierte Lindebergbedingung zu zeigen.

Wir rechnen zunächst die konditionierte Normierungsbedingung komponentenweise nach. Wie in Kapitel 5 bereits festgestellt, gilt

$$\sum_{i=1}^{n} \mathrm{E}\left(Z_{ni}^2 | \mathcal{A}_{i-1}\right) \longrightarrow \frac{\sigma^4}{1-\rho^2} \qquad \text{für } n \to \infty \qquad \text{P-stochastisch.}$$

Weiter berechnen wir mit (3.5) leicht

$$\sum_{i=1}^{n} \mathrm{E}\left(Z_{ni} A_{ni} | \mathcal{A}_{i-1}\right) = \sigma^2 \frac{1}{n} \sum_{i=1}^{n} X_{i-1} = o_{\mathrm{P}}(1)$$

und ebenso

$$\sum_{i=1}^{n} \mathrm{E}\left(Z_{ni} C_{ni}(x) | \mathcal{A}_{i-1}\right) = o_{\mathrm{P}}(1)$$

für $x \in \mathbb{R}$. Für die weiteren Einträge stellen wir fest, daß die A_{ni} und die $C_{ni}(x)$ stets unabhängig von $\mathcal{A}_{i-1}$ und identisch verteilt sind, deshalb ergibt sich

$$\sum_{i=1}^{n} \mathrm{E}\left(A_{ni}^2 | \mathcal{A}_{i-1}\right) = \sigma^2$$

und

$$\begin{aligned} \sum_{i=1}^{n} \mathrm{E}\left(C_{ni}(x) C_{ni}(y) | \mathcal{A}_{i-1}\right) &= 4\mathrm{E}\left(\left(\mathrm{h}_x^{(1)}(e_1) - \mathrm{G}(x)\right)\left(\mathrm{h}_y^{(1)}(e_1) - \mathrm{G}(y)\right)\right) \\ &= 4\mathrm{Cov}\left(\mathrm{h}_x^{(1)}(e_1), \mathrm{h}_y^{(1)}(e_1)\right) \end{aligned}$$

für $\{x, y\} \subseteq \{a, b\}$. Schließlich gilt für $x \in \{a, b\}$ wegen der Zentriertheit der e_i noch

$$\sum_{i=1}^{n} \mathrm{E}\left(A_{ni} C_{ni}(x) | \mathcal{A}_{i-1}\right) = 2\mathrm{E}\left(e_1 \mathrm{h}_x^{(1)}(e_1)\right) = \mathrm{U}_1(x) + \mathrm{U}_2(x),$$

entsprechend den Definitionen (6.23) und (6.24). Damit ist die konditionierte Normierungsbedingung bewiesen, und es bleibt die komponentenweise konditionierte Lindebergbedingung zu zeigen.

Diese gilt, wie bereits in Kapitel 5 erwähnt, für die erste Komponente des betrachteten Vektors. Weiter implizieren der Satz von Lebesgue und (2.10), daß für beliebiges $\varepsilon > 0$ die Aussage

$$\sum_{i=1}^{n} \mathrm{E}\left(A_{ni}^2 1_{\{|A_{ni}| \geq \varepsilon\}} | \mathcal{A}_{i-1}\right) = \mathrm{E}\left(e_1^2 1_{\{|e_1| \geq \varepsilon\sqrt{n}\}}\right) \longrightarrow 0 \qquad \text{für } n \to \infty$$

gilt. Genauso können wir die konditionierte Lindebergbedingung für die (sogar durch Eins beschränkte) dritte und vierte Komponente des Vektors zeigen. Damit gilt für jede Komponente die konditionierte Lindebergbedingung, und das Lemma ist bewiesen. □

Wie im Beweis zu (5.17) können wir durch Betrachtung von „gestutzten" Summen zeigen, daß auch hier die Verteilungskonvergenz

$$\begin{pmatrix} X_n \\ Z_n \\ A_n \\ C_n(b) \\ C_n(a) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} Y_\infty \\ N_1 \\ N_2 \\ N_3 \\ N_4 \end{pmatrix} \qquad \text{für } n \to \infty \tag{6.38}$$

mit der von (N_1, N_2, N_3, N_4) unabhängigen Zufallsvariablen Y_∞ gemäß (5.12) gilt.

Damit haben wir alle Hilfsmittel zum Beweis des folgenden Hauptsatzes bereitgestellt.

Satz 6.15

Unter den Voraussetzungen (2.10) und (3.13) gilt

$$\begin{aligned} &\sqrt{n}\,(\mathrm{P}\,(X_{n+2} \in \ I_{n,res}|\mathcal{A}_i) - (1-\alpha)) \\ &\xrightarrow{\mathcal{L}} \quad \{(\mathrm{g}(b) - \mathrm{g}(a))2\rho Y_\infty + (\mathrm{E}_1(b) - E_1(a))\}\,\frac{1-\rho^2}{\sigma^2} N_1 \\ &\qquad - (N_3 - N_4) \qquad \text{für } n \to \infty \end{aligned} \tag{6.39}$$

und

$$\begin{aligned} &\sqrt{n}\,(\mathrm{P}\,(X_{n+2} \in \ I_{n,res,z}|\mathcal{A}_i) - (1-\alpha)) \\ &\xrightarrow{\mathcal{L}} \quad \{(\mathrm{g}(b) - \mathrm{g}(a))2\rho Y_\infty + (\mathrm{E}_1(b) - E_1(a))\}\,\frac{1-\rho^2}{\sigma^2} N_1 - (N_3 - N_4) \\ &\qquad + \{\mathrm{U}_1(b) + \mathrm{U}_2(b) - (\mathrm{U}_1(a) + \mathrm{U}_2(a))\}\,\frac{1}{\sigma^2} N_2 \qquad \text{für } n \to \infty \end{aligned} \tag{6.40}$$

mit (N_1, N_2, N_3, N_4) wie im letzten Lemma und Y_∞, N_1, (N_2, N_3, N_4) unabhängig.

Beweis Dies folgt aus den Entwicklungen (6.35) und (6.37), der Verteilungskonvergenz (6.38), dem Stetigkeitssatz und dem Satz von Cramér-Slutzky. □

Wir haben also wie im Fall von Ein-Schritt-Prognoseintervallen eine Verteilungskonvergenzaussage für

$$\sqrt{n}\,(\mathrm{P}\,(X_{n+2} \in \mathrm{I}_{n,res}|\mathcal{A}_n) - (1-\alpha))$$

beziehungsweise für

$$\sqrt{n}\,(\mathrm{P}\,(X_{n+2} \in \mathrm{I}_{n,res,z}|\mathcal{A}_n) - (1-\alpha))$$

hergeleitet.

Wir überlegen uns jetzt, welche Verbesserung der EL-Ansatz, also die Verwendung von $\mathrm{G}_{n,res,z}$ anstelle von $\mathrm{G}_{n,res}$ erbringt. Dazu betrachten wir wie in Kaptitel 5 die Varianzen der Grenzvariablen. Bezeichnen wir mit W_1 die Grenzvariable aus (6.39) und mit W_2 die aus (6.40), so können wir leicht nachrechnen, daß

$$\mathrm{Var}(W_1) - \mathrm{Var}(W_2) = \frac{1}{\sigma^2}\left(\mathrm{U}_1(b) + \mathrm{U}_2(b) - (\mathrm{U}_1(a) + \mathrm{U}_2(a)\right)^2 \geq 0$$

gilt, was bedeutet, daß der auf dem EL-Prinzip basierende, die Zusatzinformation $\mathrm{E}(e_1) = 0$ einbeziehende, Verteilungsschätzer $\mathrm{F}_{n,res,z}$ dem kanonischen $\mathrm{F}_{n,res}$ vorzuziehen ist. Wie schon im Kapitel 5 muß aber auch hier die Einschränkung gemacht werden, daß diese Differenz für symmetrische Verteilungen gleich Null ist, so daß bei bekanntermaßen symmetrischer Fehlerverteilung F der EL-Ansatz keine Verbesserung bei der Konstruktion von Zwei-Schritt-Prognoseintervallen verspricht. Allerdings kann man wie in Kapitel 5 leicht zeigen, daß bei der Konstruktion einseitiger oberer oder unterer Prognoseschranken die oben diskutierte Verbesserung auch im Falle symmetrischer Fehlerverteilungen erhalten bleibt, so daß in diesem Fall stets das EL-Prinzip zur Schätzung von F beziehungsweise G verwendet werden sollte.

Zum Abschluß des Kapitels wollen wir noch kurz die Fälle $\rho \in (-1, 0)$ und $\rho = 0$ erörtern.

Im Fall $\rho \in (-1, 0)$ ist

$$\mathrm{P}(\rho e_1 \leq x) = 1 - \mathrm{F}(x) \qquad \text{für } x \in \mathbb{R},$$

da F stetig ist. Folglich ist in diesem Falle

$$\begin{aligned} \mathrm{G}(x) &= \mathrm{P}(\rho e_1 + e_2 \leq x) = (\mathrm{F} * (1 - \mathrm{F}(\cdot/\rho)))\,(x) \\ &= \mathrm{E}(\mathrm{F}(x - \rho e_1)) = 1 - \mathrm{E}\left(\mathrm{F}\left(\frac{x - e_1}{\rho}\right)\right), \end{aligned}$$

und es gilt

$$\mathrm{P}\left(\hat{\rho}_n \notin (-1, 0)\right) \longrightarrow 0 \qquad \text{für } n \to \infty,$$

weshalb wir stets $\hat{\rho}_n \in (-1, 0)$ annehmen können. Als Verteilungsschätzer für G definieren wir nun $\mathrm{G}_{n,res} = \mathrm{F}_{n,res} * (1 - \mathrm{F}_{n,res}(\cdot/\hat{\rho}_n))$ beziehungsweise $\mathrm{G}_{n,res,z} = \mathrm{F}_{n,res,z} * (1 - \mathrm{F}_{n,res,z}(\cdot/\hat{\rho}_n))$. Für die mit den Quantilen dieser Verteilungsfunktionen definierten Zwei-Schritt-Prognoseintervalle gelten sinngemäß alle Ergebnisse, die wir im Fall $\rho \in (0, 1)$ explizit bewiesen haben.

Im Fall $\rho = 0$ gilt $\mathrm{G} = \mathrm{F}$ und $X_n = e_n$ für $n \geq 1$. Weiter ist wegen der stetigen Verteiltheit der e_n die Menge $\{\hat{\rho}_n = 0\}$ eine P-Nullmenge, das heißt

$$\mathrm{P}\left(\{\hat{\rho}_n < 0\} \cup \{\hat{\rho}_n > 0\}\right) = 1.$$

Deshalb ist die Definition

$$\mathrm{H}_n(x) = \mathrm{F}_n(x/\hat{\rho}_n)1_{\{\hat{\rho}_n>0\}} + (1-\mathrm{F}_n(x/\hat{\rho}_n))1_{\{\hat{\rho}_n<0\}} \qquad \text{für } x \in \mathbb{R}$$

sinnvoll. Weiter sei $\mathrm{H}_{n,res}$ entsprechend mit $\mathrm{F}_{n,res}$ anstelle von F_n definiert und

$$\mathrm{G}_{n,res}(x) = (\mathrm{F}_{n,res} * \mathrm{H}_{n,res})(x) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1_{\{\hat{e}_i+\hat{\rho}_n\hat{e}_j\leq x\}} \qquad \text{für } x \in \mathbb{R}.$$

Wegen (6.1) und (3.14) gilt

$$\sup_{x\in\mathbb{R}} |\mathrm{G}_{n,res}(x) - (\mathrm{F}_n * \mathrm{H}_n)(x)| = o_{\mathrm{P}}\left(n^{-1/2}\right), \tag{6.41}$$

und wegen

$$\sup_{x\in\mathbb{R}} |\mathrm{F}_n * \mathrm{H}_n(x) - \mathrm{F}(x)| = o_{\mathrm{P}}(1)$$

erhalten wir

$$\sup_{x\in\mathbb{R}} |\mathrm{G}_{n,res}(x) - \mathrm{F}(x)| = o_{\mathrm{P}}(1),$$

sowie mit (5.4) und (3.13) weiter

$$\mathrm{G}_{n,res}^{-1}(u) \longrightarrow \mathrm{F}^{-1}(u) \qquad \text{für } n\to\infty$$

für beliebiges $u \in (0,1)$.

Dies zeigt zusammen mit

$$\hat{\rho}_n^2 X_n = \hat{\rho}_n^2 e_n = O_{\mathrm{P}}\left(n^{-1}\right),$$

daß

$$\begin{aligned} &\mathrm{P}\left(X_{n+2} \in \left[\ \hat{\rho}_n^2 X_n + \mathrm{G}_{n,res}^{-1}(v), \hat{\rho}_n^2 X_n + \mathrm{G}_{n,res}^{-1}(w)\ \right] |\mathcal{A}_n\right) \\ &= \ \mathrm{P}\left(e_{n+2} \in \left[\ \hat{\rho}_n^2 e_n + \mathrm{G}_{n,res}^{-1}(v), \hat{\rho}_n^2 e_n + \mathrm{G}_{n,res}^{-1}(w)\ \right] |\mathcal{A}_n\right) \longrightarrow 1-\alpha \end{aligned}$$

für $n\to\infty$ gilt. Also konvergieren die bedingten Überdeckungswahrscheinlichkeiten der mit $\mathrm{G}_{n,res}$ konstruierten Zwei-Schritt-Prognoseintervalle gegen das nominelle Niveau.

Auch über die Konvergenzgeschwindigkeit können wir eine zum Fall $\rho \in (0,1)$ analoge Verteilungskonvergenzaussage machen.

Dazu zeigen wir zunächst, daß für beliebiges $C > 0$ die Beziehung

$$\sup_{\substack{|u|\leq C\\ x\in\mathbb{R}}} \left| \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1_{\{e_i+un^{-1/2}e_j\leq x\}} - \mathrm{F}(x) - (\mathrm{F}_n(x) - \mathrm{F}(x)) \right| = o_{\mathrm{P}}\left(n^{-1/2}\right) \tag{6.42}$$

gilt. Der Beweis hierfür erfolgt zum größten Teil mit den selben Methoden wie im Fall $\rho \in (0,1)$. Allerdings stoßen wir auf das Problem, daß wir in

$$2\mathrm{h}_{u,x}^{(1)}(s) = \mathrm{F}(x - un^{-1/2}s) + \mathrm{E}(1_{\{s+un^{-1/2}e_1 \leq x\}}) \qquad \text{für } s \in \mathbb{R}$$

den zweiten Summanden offensichtlich nicht mehr geschlossen als Funktionswert der Verteilungsfunktion F schreiben können, was bei $\rho \in (0,1)$ möglich war. An dieser Stelle verwenden wir stattdessen, daß

$$\mathrm{E}(1_{\{s+un^{-1/2}e_1 \leq x\}}) = \int_{\mathbb{R}} 1_{\{s \leq x - un^{-1/2}t\}} \mathrm{F}(dt)$$

und somit

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{E}(1_{\{s+un^{-1/2}e_1 \leq x\}}) \circ e_i &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\mathbb{R}} 1_{\{e_i \leq x - un^{-1/2}t\}} \mathrm{F}(dt) \\
&= \int_{\mathbb{R}} \sqrt{n} \mathrm{F}_n(x - un^{-1/2}t) \mathrm{F}(dt)
\end{aligned}$$

gilt. Hieraus können wir

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{E}(1_{\{s+un^{-1/2}e_1 \leq x\}}) \circ e_i - \sqrt{n} F_n(x) \\
&= \int_{\mathbb{R}} \sqrt{n} \left(\mathrm{F}_n(x - un^{-1/2}t) - \mathrm{F}(x - un^{-1/2}t) - (\mathrm{F}_n(x) - \mathrm{F}(x)) \right) \mathrm{F}(dt) \\
&\quad + \int_{\mathbb{R}} \sqrt{n} \left(\mathrm{F}(x - un^{-1/2}t) - \mathrm{F}(x) \right) \mathrm{F}(dt) \\
&= R_{1n}(u,x) + R_{2n}(u,x)
\end{aligned}$$

folgern. Wegen der bekannten C-Straffheit des empirischen Prozesses gibt es insbesondere zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß

$$\limsup_{n \to \infty} \mathrm{P} \left(\sqrt{n} \sup_{|k-l| \leq \delta} |\mathrm{F}_n(k) - \mathrm{F}(k) - (\mathrm{F}_n(l) - \mathrm{F}(l))| \geq \frac{\varepsilon}{2} \right) \leq \varepsilon$$

gilt. Für ein solches δ schätzen wir

$$\begin{aligned}
\sup_{\substack{|u| \leq C \\ x \in \mathbb{R}}} |R_{1n}(u,x)| &\leq \int_{\mathbb{R}} \sqrt{n} \sup_{|k-l| \leq Cn^{-1/2}|t|} |\mathrm{F}_n(k) - \mathrm{F}(k) - (\mathrm{F}_n(l) - \mathrm{F}(l))| \, \mathrm{F}(dt) \\
&\leq \int_{\mathbb{R}} \sqrt{n} \sup_{|k-l| \leq \delta} |\mathrm{F}_n(k) - \mathrm{F}(k) - (\mathrm{F}_n(l) - \mathrm{F}(l))| \, 1_{\left\{|t| \leq \frac{\delta\sqrt{n}}{C}\right\}} \mathrm{F}(dt) \\
&\quad + 2 \int_{\mathbb{R}} \sqrt{n} 1_{\left\{|t| > \frac{\delta\sqrt{n}}{C}\right\}} \mathrm{F}(dt) \\
&\leq \sqrt{n} \sup_{|k-l| \leq \delta} |\mathrm{F}_n(k) - \mathrm{F}(k) - (\mathrm{F}_n(l) - \mathrm{F}(l))| \\
&\quad + \frac{2C}{\delta} \mathrm{E} \left(|e_1| 1_{\left\{|e_1| > \frac{\delta\sqrt{n}}{C}\right\}} \right)
\end{aligned}$$

ab, und der zweite Summand auf der rechten Seite der Ungleichung konvergiert auf Grund des Satzes von Lebesgue in n gegen Null. Somit gilt

$$\limsup_{n\to\infty} \mathrm{P}\left(\sup_{\substack{|u|\le C\\ x\in\mathbb{R}}} |R_{1n}(u,x)| \ge \varepsilon \right)$$
$$\le \limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n} \sup_{|k-l|\le\delta} |\mathrm{F}_n(k) - \mathrm{F}(k) - (\mathrm{F}_n(l) - \mathrm{F}(l))| \ge \frac{\varepsilon}{2} \right) \le \varepsilon,$$

was die gewünschte stochastische Konvergenz für $R_{1n}(u,x)$ beweist.

Ebenso wie im Fall $\rho \in (0,1)$ zeigt man

$$\sup_{\substack{|u|\le C\\ x\in\mathbb{R}}} |R_{2n}(u,x)| = o_{\mathrm{P}}(1)$$

und die übrigen noch nötigen Schritte zum Beweis von (6.42). Hieraus wiederum folgt mit (6.41) sofort

$$\mathrm{G}_{n,res}(x) - \mathrm{F}(x) = \mathrm{F}_n(x) - \mathrm{F}(x) + R_n(x)$$

mit

$$\sup_{x\in\mathbb{R}} |R_n(x)| = o_P\left(n^{-1/2}\right)$$

und wegen der C-Straffheit des empirischen Prozesses auch

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathrm{P}\left(\sqrt{n} \sup_{|y-x|\le\delta} |\mathrm{G}_{n,res}(y) - \mathrm{F}(y) - (\mathrm{G}_{n,res}(x) - \mathrm{F}(x))| \ge \varepsilon \right) = 0$$

für beliebiges $\varepsilon > 0$. Dies impliziert

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+2} \in I_{n,res}|\mathcal{A}_n\right) - (1-\alpha)\right) = -\sqrt{n}(\mathrm{F}_n(b) - w) + \sqrt{n}(\mathrm{F}_n(a) - v) + o_{\mathrm{P}}(1),$$

also

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+2} \in I_{n,res}|\mathcal{A}_n\right) - (1-\alpha)\right) \xrightarrow{\mathcal{L}} -N_1 + N_2 \qquad \text{für } n \to \infty$$

mit

$$\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \stackrel{\mathcal{L}}{=} \mathcal{N}_2\left(0, \Gamma_1\right)$$

und

$$\Gamma_1 = \begin{pmatrix} vw & v^2 \\ v^2 & vw \end{pmatrix}.$$

Letzteres wurde in (5.15) gezeigt.

Ersetzt man nun überall die empirische Verteilungsfunktion der Residuen $\mathrm{F}_{n,res}$ durch ihr EL-Analogon $\mathrm{F}_{n,res,z}$, so folgt ebenso wie oben

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+2} \in I_{n,res,z} | \mathcal{A}_n\right) - (1-\alpha)\right) \xrightarrow{\mathcal{L}} -N_3 + N_4 \qquad \text{für } n \to \infty$$

mit

$$\begin{pmatrix} N_3 \\ N_4 \end{pmatrix} \stackrel{\mathcal{L}}{=} \mathcal{N}_2\left(0, \Gamma_2\right)$$

und

$$\Gamma_2 = \begin{pmatrix} vw - \frac{\mathrm{U}(a)^2}{\sigma^2} & v^2 - \frac{\mathrm{U}(a)\mathrm{U}(b)}{\sigma^2} \\ v^2 - \frac{\mathrm{U}(a)\mathrm{U}(b)}{\sigma^2} & vw - \frac{\mathrm{U}(b)^2}{\sigma^2} \end{pmatrix}.$$

Ein Vergleich der Grenzvarianzen ergibt das schon bekannte Bild. Es gilt

$$\mathrm{Var}(N_1 - N_2) - \mathrm{Var}(N_3 - N_4) = \frac{(\mathrm{U}(b) - \mathrm{U}(a))^2}{\sigma^2} \geq 0,$$

also ist auch hier der EL-Schätzer dem kanonischen vorzuziehen, falls $\mathrm{E}(e_1) = 0$ bekannt ist.

In der Praxis weiß man auch unter der Modellannahme, daß $(X_n)_{n\geq 0}$ ein stabiler AR(1)-Prozeß ist, im allgemeinen nicht, ob $\rho \in (0,1)$, $\rho \in (-1,0)$ oder $\rho = 0$ zutreffend ist. Dann sollte

$$\mathrm{G}_{n,res} = \mathrm{F}_{n,res} * \mathrm{H}_{n,res}$$

definiert werden, da zum Beispiel im Fall $\rho \in (0,1)$ dann

$$\mathrm{P}\left(\mathrm{H}_{n,res}(x) \neq \mathrm{F}_{n,res}(x) \text{ für ein } x \in \mathbb{R}\right) \leq \mathrm{P}\left(\hat{\rho}_n \leq 0\right) = o(1)$$

gilt. Analoges gilt für die mit dem EL-Verfahren erhaltenen Schätzer.

Somit liefert das EL-Varfahren bei stabilen AR(1)-Prozessen mit zentrierten Fehlern stets Zwei-Schritt-Prognoseintervalle, bei denen die Konvergenz der bedingten Überdeckungswahrscheinlichkeiten gegen das nominelle Niveau besser, das heißt mit einer nichtgrößeren Grenzvarianz erfolgt.

Kapitel 7

Simulationsstudien

Im letzten Kapitel der Arbeit wollen wir die gewonnenen Ergebnisse empirisch verdeutlichen. Dabei werden zu ausgewählten Stichprobenunmfängen, Verteilungen, Niveaus und – im AR(1)-Modell – ρ-Werten Computersimulationen durchgeführt und die auf dem EL-Prinzip basierenden Ergebnisse mit den auf der klassischen Schätzung beruhenden verglichen.

7.1 Anpassungstests

7.1.1 Anpassungstests bei unabhängig und identisch verteilten Zufallsvariablen

In diesem Abschnitt werden die Schätzer F_n und $\mathrm{F}_{n,z}$ im Hinblick auf Anpassungstests im Modell unabhängig und identisch verteilter, zentrierter Zufallsvariablen verglichen. Wir testen auf die Hypothese (2.2) also auf

$$\mathrm{H}_0 : \mathrm{F} \in \mathcal{F} \qquad \text{gegen} \qquad \mathrm{H}_1 : \mathrm{F} \notin \mathcal{F}$$

mit

$$\mathcal{F} = \left\{ \mathrm{F}(\cdot, \vartheta) | \vartheta \in \Theta \subset \mathrm{I\!R}^d \right\} \tag{2.3}$$

Untersuchte Verteilungklassen sind die Klasse der zentrierten Normalverteilungen (kurz $\mathcal{F}_1$) und die Skalenmodelle der Doppelexponentialverteilung (kurz $\mathcal{F}_2$) sowie der t_3-Verteilung (kurz $\mathcal{F}_3$), der t-Verteilung mit drei Freiheitsgraden. Dagegen wurden die wahren Verteilungen Standarnormalverteilung (kurz $\mathcal{N}(0,1)$), Doppelexponentialverteilung (kurz Dexp), t_3-Verteilung (kurz t_3) und Standard-Cauchy-Verteilung (kurz $C(0,1)$) getestet.

Entsprechende Zufallsvariablen konnten im Fall der Doppelexponentialverteilung und der Cauchyverteilung mittels Quantiltransformation erzeugt werden. Für die

Simulation normalverteilter Variablen wurde die Box-Müller-Tranformation (vgl. Press et al. 1992, S.289), für die t_3-verteilter ein Algorithmus aus Devroye (1986, S.449) verwendet. Der benutzte Schätzer für den Verteilungsparameter ist stets der unter der Hypothese errechnete ML-Schätzer.

In der Simulation werden 1000 Werte der klassischen Kolmogorow-Smirnow-Statistik

$$\sup_{x\in\mathbb{R}} \left|\sqrt{n}(\mathrm{F}_n(x) - \mathrm{F}(x, \hat{\vartheta}_n))\right| \tag{KL}$$

beziehungsweise der EL-basierten

$$\sup_{x\in\mathbb{R}} \left|\sqrt{n}(\mathrm{F}_{n,z}(x) - \mathrm{F}(x, \hat{\vartheta}_n))\right| \tag{EL}$$

mit den auf Grund von jeweils 2000 Bootstrapstichproben berechneten kritischen Werten verglichen. In den folgenden Tabellen sind die Verwerfungshäufigkeiten zusammengefaßt.

Tab. 1	Hypothese	$\mathcal{F}_1$		$\mathcal{F}_2$		$\mathcal{F}_3$	
wahre Verteilung		$\mathcal{N}(0,1)$		Dexp		t_3	
Teststatistik		KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.010	0.015	0.014	0.009	0.009	0.013
	60	0.008	0.011	0.007	0.008	0.017	0.010
	100	0.007	0.011	0.012	0.015	0.012	0.010
	500	0.010	0.010	0.009	0.007	0.013	0.007
$\alpha = 0.05$	$n = 20$	0.051	0.053	0.048	0.053	0.055	0.059
	60	0.045	0.048	0.041	0.040	0.063	0.048
	100	0.048	0.044	0.049	0.047	0.062	0.053
	500	0.049	0.051	0.065	0.043	0.050	0.046
$\alpha = 0.1$	$n = 20$	0.100	0.097	0.096	0.090	0.110	0.108
	60	0.100	0.098	0.092	0.098	0.103	0.092
	100	0.099	0.097	0.089	0.090	0.110	0.101
	500	0.091	0.099	0.108	0.096	0.095	0.088

Tab. 2	Hypothese	$\mathcal{F}_1$					
wahre Verteilung		Dexp		t_3		$C(0,1)$	
Teststatistik		KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.014	0.056	0.017	0.105	0.331	0.664
	60	0.035	0.244	0.076	0.314	0.919	0.991
	100	0.092	0.438	0.185	0.534	0.996	1.000
$\alpha = 0.05$	$n = 20$	0.062	0.173	0.086	0.222	0.553	0.797
	60	0.180	0.473	0.226	0.483	0.972	1.000
	100	0.282	0.674	0.388	0.722	0.999	1.000
$\alpha = 0.1$	$n = 20$	0.131	0.269	0.152	0.314	0.660	0.840
	60	0.320	0.595	0.349	0.602	0.988	1.000
	100	0.441	0.795	0.544	0.802	1.000	1.000

Tab. 3	Hypothese	$\mathcal{F}_2$					
wahre Verteilung		$\mathcal{N}(0,1)$		t_3		$C(0,1)$	
Teststatistik		KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.023	0.007	0.017	0.017	0.103	0.298
	60	0.031	0.024	0.020	0.026	0.491	0.751
	100	0.057	0.045	0.017	0.031	0.729	0.925
	500	0.356	0.800	–	–	–	–
$\alpha = 0.05$	$n = 20$	0.084	0.049	0.058	0.083	0.243	0.453
	60	0.122	0.118	0.072	0.096	0.688	0.843
	100	0.189	0.228	0.080	0.092	0.868	0.974
	500	0.747	0.980	–	–	–	–
$\alpha = 0.1$	$n = 20$	0.148	0.108	0.114	0.141	0.349	0.542
	60	0.212	0.233	0.132	0.170	0.770	0.909
	100	0.320	0.395	0.134	0.177	0.932	0.987
	500	0.921	0.994	–	–	–	–

Tab. 4	Hypothese	$\mathcal{F}_3$					
wahre Verteilung		$\mathcal{N}(0,1)$		Dexp		$C(0,1)$	
Teststatistik		KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.018	0.004	0.009	0.010	0.012	0.072
	60	0.021	0.001	0.016	0.011	0.031	0.172
	100	0.019	0.003	0.007	0.009	0.045	0.248
	500	0.053	0.027	0.033	0.022	–	–
	1000	0.140	0.129	0.053	0.047	–	–
$\alpha = 0.05$	$n = 20$	0.064	0.026	0.061	0.040	0.074	0.223
	60	0.074	0.023	0.055	0.048	0.133	0.379
	100	0.087	0.028	0.043	0.043	0.206	0.508
	500	0.214	0.189	0.125	0.118	–	–
	1000	0.405	0.549	0.175	0.223	–	–
$\alpha = 0.1$	$n = 20$	0.131	0.066	0.117	0.082	0.144	0.338
	60	0.138	0.062	0.101	0.092	0.248	0.526
	100	0.148	0.078	0.099	0.085	0.341	0.665
	500	0.346	0.367	0.200	0.233	–	–
	1000	0.614	0.813	0.291	0.376	–	–

Wie die Simulationsergebnisse zeigen, halten sowohl der auf F_n als auch der auf $\mathrm{F}_{n,z}$ beruhende Test auf der Hypothese das nominelle Niveau approximativ gut ein (Tab. 1).

Vergleicht man die Güte der Tests, so ergibt sich das folgende Bild: Hat die wahre Verteilung schwerere Tails als die Verteilungen der Hypothesenklasse, so wird dies von beiden Tests gut erkannt. Die Güte ist dabei umso größer, je stärker sich wahre Verteilung und Verteilungshypothese im Tailverhalten unterscheiden. Der EL-basierte Test hat allerdings eine teilweise wesentlich größere Gütefunktion als der klassische (vgl. Tab. 2; Tab. 3, Spalten 5-8; Tab. 4, Spalten 7+8).

Hat die wahre Verteilung jedoch leichtere Tails als die Verteilungen der Hypothesenklasse, so ist die Güte bei beiden Testverfahren erheblich geringer. Je weniger sich die Verteilungen der Hypothesenklasse und die wahre Verteilung im Tailverhalten unterscheiden, desto seltener wird die falsche Hypothese verworfen. Bei kleinen Stichprobenumfängen ist die Güte des EL-Tests kleiner als die der klassischen und liegt sogar gelegentlich unter dem nominellen Niveau. Bei kleinen Stichprobenumfängen ist der klassische Test also dem EL-Test überlegen. Allerdings zeichnet sich ab, daß der EL-Test den klassischen mit wachsendem Stichprobenumfang „überholt" (vgl. Tab. 3, Spalten 3+4; Tab. 4, Spalten 3-6).

7.1.2 Anpassungstests im AR(1)-Modell

Für entsprechende Tests im AR(1)-Modell ist natürlich zunächst der Autoregressionsparameter ρ zu schätzen. Um die Ergebnisse aus Kapitel 3 anwenden zu können, muß der Schätzer (3.2) erfüllen. Wir haben bereits den Ansatz von Datta und Sriram (1997) zur Modifikation des Kleinst-Quadrate-Schätzers $\tilde{\rho}_n$ diskutiert. Für die Anwendung in Computersimulationen hat sich die folgende Modifikation als vorteilhaft erwiesen. Wir definieren

$$\hat{\rho}_n = \begin{cases} -1 & \text{für } \tilde{\rho}_n \in (-1 - n^{-0.01}, -1 + n^{-0.01}) \\ 1 & \text{für } \tilde{\rho}_n \in (1 - n^{-0.01}, 1 + n^{-0.01}) \\ \tilde{\rho}_n & \text{sonst.} \end{cases}$$

Im Vergleich mit der in Kapitel 3 betrachteten Version sind hier die Intervalle, auf denen der Kleinst-Quadrate-Schätzer verändert wird, kleiner gewählt, was die Häufigkeit einer inadäquaten Schätzung im Fall $|\rho| \neq 1$ reduziert. Trotzdem sind die Ergebnisse, wie wir sehen werden, auch im Fall $|\rho| = 1$ annehmbar, und natürlich erfüllt der so definierte Schätzer auch die Voraussetzung (3.2).

Simuliert wurden wiederum 1000 Werte der Teststatistiken

$$\sup_{x \in \mathbb{R}} \left| \sqrt{n}(\mathrm{F}_{n,res}(x) - \mathrm{F}(x, \hat{\vartheta}_{n,res})) \right| \tag{KL}$$

und

$$\sup_{x \in \mathbb{R}} \left| \sqrt{n}(\mathrm{F}_{n,res,z}(x) - \mathrm{F}(x, \hat{\vartheta}_{n,res})) \right| \tag{EL}$$

mit zu $\hat{\rho}_n$ gebildeten Residuen sowie jeweils 2000 Bootstrapkopien zur Berechnung der kritischen Werte. Als Werte für den Autoregressionsparameter wurden $\rho = 0.0, 0.5, 1.0, 1.05$ verwendet. Die Abweichungen der Simulationsergebnisse vom Fall unabhängig und identisch verteilter Variablen sind für jede der betrachteten Hypothesenklassen ähnlich, so daß wir uns aus Platzgründen auf die exemplarische Darstellung der Ergebnisse zur Hypothese $\mathcal{F}_1$ (zentrierte Normalverteilungen) beschränken. Die folgenden Tabellen geben die Verwerfungshäufigkeiten wieder.

Tab. 5		Hypothese: $\mathcal{F}_1$, wahre Verteilung: $\mathcal{N}(0,1)$							
		$\rho = 0.0$		$\rho = 0.5$		$\rho = 1.0$		$\rho = 1.05$	
Teststatistik		KL	EL	KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.005	0.009	0.002	0.008	0.001	0.009	0.006	0.006
	60	0.003	0.008	0.007	0.014	0.005	0.019	0.003	0.009
	100	0.004	0.014	0.010	0.007	0.009	0.008	0.003	0.008
	500	0.013	0.008	0.012	0.015	0.002	0.012	0.006	0.010
$\alpha = 0.05$	$n = 20$	0.035	0.047	0.034	0.041	0.031	0.043	0.047	0.043
	60	0.028	0.042	0.043	0.055	0.038	0.046	0.022	0.051
	100	0.038	0.048	0.051	0.052	0.041	0.048	0.027	0.046
	500	0.049	0.055	0.051	0.047	0.034	0.052	0.037	0.057
$\alpha = 0.1$	$n = 20$	0.082	0.078	0.078	0.091	0.104	0.102	0.096	0.091
	60	0.087	0.081	0.093	0.095	0.088	0.106	0.056	0.103
	100	0.093	0.094	0.098	0.102	0.090	0.097	0.070	0.096
	500	0.091	0.100	0.101	0.093	0.098	0.105	0.080	0.104

Tab. 6		Hypothese: $\mathcal{F}_1$, wahre Verteilung: Dexp							
		$\rho = 0.0$		$\rho = 0.5$		$\rho = 1.0$		$\rho = 1.05$	
Teststatistik		KL	EL	KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.002	0.050	0.003	0.050	0.004	0.043	0.021	0.062
	60	0.029	0.223	0.032	0.240	0.028	0.228	0.021	0.236
	100	0.074	0.405	0.076	0.407	0.084	0.417	0.054	0.441
$\alpha = 0.05$	$n = 20$	0.043	0.159	0.044	0.172	0.046	0.134	0.090	0.187
	60	0.155	0.430	0.158	0.456	0.153	0.437	0.148	0.469
	100	0.262	0.637	0.269	0.668	0.303	0.673	0.241	0.678
$\alpha = 0.1$	$n = 20$	0.112	0.248	0.101	0.255	0.113	0.232	0.193	0.293
	60	0.284	0.572	0.267	0.575	0.313	0.575	0.285	0.593
	100	0.420	0.756	0.446	0.778	0.488	0.775	0.409	0.791

Tab. 7		Hypothese: $\mathcal{F}_1$, wahre Verteilung: t_3							
		$\rho = 0.0$		$\rho = 0.5$		$\rho = 1.0$		$\rho = 1.05$	
Teststatistik		KL	EL	KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.010	0.094	0.005	0.093	0.014	0.091	0.029	0.078
	60	0.069	0.296	0.089	0.325	0.095	0.336	0.081	0.338
	100	0.171	0.529	0.170	0.538	0.183	0.539	0.161	0.527
$\alpha = 0.05$	$n = 20$	0.066	0.203	0.055	0.187	0.081	0.203	0.099	0.199
	60	0.196	0.474	0.243	0.496	0.249	0.523	0.238	0.494
	100	0.392	0.712	0.379	0.726	0.408	0.713	0.359	0.703
$\alpha = 0.1$	$n = 20$	0.139	0.298	0.133	0.278	0.146	0.294	0.192	0.279
	60	0.318	0.597	0.355	0.587	0.369	0.620	0.345	0.594
	100	0.524	0.798	0.514	0.792	0.558	0.797	0.499	0.797

Tab. 8		Hypothese: $\mathcal{F}_1$, wahre Verteilung: $C(0,1)$							
		$\rho = 0.0$		$\rho = 0.5$		$\rho = 1.0$		$\rho = 1.05$	
Teststatistik		KL	EL	KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.303	0.627	0.297	0.639	0.298	0.639	0.313	0.628
	60	0.906	0.994	0.924	0.992	0.913	0.991	0.921	0.994
	100	0.993	1.000	0.993	1.000	0.998	1.000	0.993	0.999
$\alpha = 0.05$	$n = 20$	0.513	0.756	0.494	0.771	0.520	0.778	0.542	0.756
	60	0.969	0.999	0.970	0.996	0.962	0.995	0.975	0.998
	100	0.999	1.000	0.998	1.000	0.999	1.000	0.998	0.999
$\alpha = 0.1$	$n = 20$	0.624	0.806	0.621	0.830	0.657	0.835	0.668	0.814
	60	0.982	1.000	0.984	0.997	0.983	0.997	0.987	0.999
	100	0.999	1.000	1.000	1.000	1.000	1.000	0.999	0.999

In allen Fällen ist der Einfluß von ρ eher gering. Auch die Daten im Fall $\rho = 1$ fügen sich ins Gesamtbild ein.

Unter der Hypothese findet die Konvergenz gegen das nominelle Niveau langsamer statt als bei unabhängig und identisch verteilten Zufallsvariablen, was vermutlich auf den zusätzlichen zufälligen Einfluß aus der Schätzung des Autoregressionsparameters zurückzuführen ist. Dabei stabilisieren sich die Verwerfungshäufigkeiten des EL-Tests überwiegend rascher (vgl. Tab. 5).

Unter den verschiedenen Alternativen ergibt sich das schon bekannte Bild. Beide Tests verwerfen umso häufiger, je stärker das Tailverhalten der wahren Verteilung von dem einer Normalverteilung abweicht. Dabei ist die Güte bei beiden Tests anscheinend etwas geringer als im Fall unabhängig und identisch verteilter Daten. Die Vorteilhaftigkeit des EL-Tests gegenüber dem klassischen ist aber wieder sehr deutlich ersichtlich (vgl. Tab. 6-8).

Zusammenfassend können wir sagen, daß der EL-basierte Test sowohl bei unabhängig und identisch verteilten zentrierten Variablen als auch im AR(1)-Modell bei Hypothesenklassen mit leichten Tails wegen der größeren Güte klar vorteilhaft ist. Hat die Hypothesenklasse schwere Tails, so ist der EL-Test nur dann günstiger als der klassische, falls der Stichprobenumfang sehr groß ist.

7.2 Prognosen

Im letzten Abschnitt dieses Kapitels wollen wir die über Prognoseintervalle und Prognoseschranken gewonnenen Ergebnisse empirisch untersuchen. Dazu betrachten wir AR(1)-Modelle zu den Parameterwerten $\rho = 0.0, 0.3, 0.5, 0.95$ mit unabhängig und identisch standardnormalverteilten sowie t_3-verteilten Fehlern. Die Simulation der Zufallsvariablen erfolgt dabei mithilfe der obengenannten Algorithmen und die

Schätzung von ρ durch den Kleinst-Quadrate-Schätzer $\hat{\rho}_n$. Zu je 1000 Ein-Schritt-Prognoseintervallen der Form

$$I_{n,res} = \left[\hat{\rho}_n X_n + \mathrm{F}_{n,res}^{-1}\left(\frac{\alpha}{2}\right), \hat{\rho}_n X_n + \mathrm{F}_{n,res}^{-1}\left(1-\frac{\alpha}{2}\right) \right] \tag{KL}$$

beziehungsweise

$$I_{n,res,z} = \left[\hat{\rho}_n X_n + \mathrm{F}_{n,res,z}^{-1}\left(\frac{\alpha}{2}\right), \hat{\rho}_n X_n + \mathrm{F}_{n,res,z}^{-1}\left(1-\frac{\alpha}{2}\right) \right] \tag{EL}$$

überprüfen wir durch Simulation von jeweils 10000 Prognosewerten die (unbedingte) Überdeckung. Die mittleren Überdeckungshäufigkeiten sind in den beiden folgenden Tabellen zusammengefaßt.

Tabelle 9		Fehlerverteilung: $\mathcal{N}(0,1)$							
		$\rho = 0.0$		$\rho = 0.3$		$\rho = 0.5$		$\rho = 0.95$	
Intervalltyp		KL	EL	KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.894	0.894	0.891	0.891	0.889	0.889	0.885	0.885
	100	0.980	0.980	0.978	0.978	0.980	0.980	0.979	0.979
	1000	0.989	0.989	0.989	0.989	0.989	0.989	0.989	0.989
$\alpha = 0.05$	$n = 20$	0.894	0.891	0.891	0.888	0.889	0.887	0.885	0.884
	100	0.940	0.942	0.937	0.939	0.939	0.941	0.938	0.940
	1000	0.949	0.949	0.949	0.950	0.949	0.949	0.949	0.949
$\alpha = 0.1$	$n = 20$	0.840	0.853	0.839	0.849	0.841	0.850	0.837	0.850
	100	0.891	0.894	0.886	0.889	0.889	0.891	0.888	0.889
	1000	0.899	0.899	0.899	0.900	0.899	0.899	0.899	0.899

Tabelle 10		Fehlerverteilung: t_3							
		$\rho = 0.0$		$\rho = 0.3$		$\rho = 0.5$		$\rho = 0.95$	
Intervalltyp		KL	EL	KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.894	0.894	0.897	0.897	0.899	0.899	0.888	0.887
	100	0.979	0.978	0.980	0.979	0.979	0.979	0.980	0.979
	1000	0.989	0.989	0.989	0.989	0.989	0.989	0.989	0.989
$\alpha = 0.05$	$n = 20$	0.894	0.887	0.897	0.890	0.899	0.891	0.888	0.879
	100	0.939	0.942	0.940	0.942	0.938	0.941	0.940	0.942
	1000	0.949	0.949	0.949	0.949	0.949	0.949	0.949	0.949
$\alpha = 0.1$	$n = 20$	0.844	0.852	0.845	0.852	0.848	0.853	0.834	0.846
	100	0.889	0.892	0.890	0.893	0.890	0.892	0.889	0.891
	1000	0.899	0.899	0.898	0.899	0.899	0.899	0.899	0.899

Wie deutlich zu sehen ist, sind die Unterschiede zwischen den beiden Intervalltypen äußerst gering. Auch ein Einfluß des Autoregressionsparameters auf die Überdeckung ist nicht ersichtlich. Selbst das Tailverhalten der Fehlerverteilungen scheint keinen Unterschied zu machen. Die Überdeckung ist für große Stichprobenumfänge sehr gut, bei kleinen Niveaus und kleinen Stichprobenumfängen ist sie jedoch stets zu klein. Dies ist nicht weiter verwunderlich, da die Intervallgrenzen wesentlich durch Quantile diskreter Verteilungen also durch Ordnungsstatistiken der Residuen determiniert sind. Ist zum Beispiel $\alpha = 0.01$ und $n = 20$, so gilt

$$\mathrm{F}_{n,res}\left(\min_{1\le i\le 20} \hat{e}_i\right) = \frac{1}{20} \qquad \text{und} \qquad \mathrm{F}_{n,res,z}\left(\min_{1\le i\le 20} \hat{e}_i\right) \approx \frac{1}{20},$$

also wird in diesem Fall nahezu immer die erste Ordnungsstatistik der Residuen zur Definition der Prognoseintervalle herangezogen. Das bedeutet aber eine im Mittel zu große Schätzung des 0.005-Quantils der Fehlerverteilung. Analog wird das 0.995-Quantil im Mittel zu klein geschätzt, was zu kurze Intervalle und damit zu geringe Überdeckungshäufigkeiten zur Folge hat. Aus diesen Überlegungen können wir folgern, daß für eine adäquate Quantilschätzung ein Stichprobenumfang von mindestens

$$n \ge \frac{2}{\alpha}$$

vorliegen sollte.

Weiter wurden zu jedem simulierten Intervall der Wert

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in I_{n,res} | \mathcal{A}_n\right) - (1-\alpha)\right) \tag{KL}$$

beziehungsweise

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \in I_{n,res,z} | \mathcal{A}_n\right) - (1-\alpha)\right) \tag{EL}$$

berechnet. Durch Bilden des Mittelwertes (MW) und der Stichprobenvarianz (SV) dieser Daten erhalten wir Näherungen für Erwartungswert und Varianz der Grenzvariablen aus Kapitel 5. Diese sind bei bekanntermaßen symmetrischer Fehlerverteilung bekannt (vgl. Satz 5.11). So ist der Erwartungswert der Grenzvariablen immer Null, während die Grenzvarianz stets den Wert $\alpha(1-\alpha)$ hat. In den folgenden Tabellen wird die Annäherung der empirischen Größen (MW) und (SV) gegen die jeweiligen Grenzwerte aufgezeigt. Da der Einfluß von ρ wieder vernachlässigbar ist, beschränken wir uns auf den Fall $\rho = 0.0$.

Tabelle 11		Fehlerverteilung: $\mathcal{N}(0,1)$				
		MW		SV		wahre
Intervalltyp		KL	EL	KL	EL	Grenzvarianz
$\alpha = 0.01$	$n = 20$	-0.429	-0.429	0.0996	0.0996	
	100	-0.100	-0.100	0.0200	0.0200	0.0099
	1000	-0.034	-0.031	0.0112	0.0110	
$\alpha = 0.05$	$n = 20$	-0.250	-0.263	0.0996	0.1100	
	100	-0.097	-0.079	0.0587	0.0603	0.0475
	1000	-0.036	-0.027	0.0465	0.0456	
$\alpha = 0.1$	$n = 20$	-0.267	-0.212	0.1432	0.1353	
	100	-0.090	-0.062	0.1002	0.0925	0.09
	1000	-0.039	-0.027	0.0906	0.0902	

Tabelle 12		Fehlerverteilung: t_3				
		MW		SV		wahre
Intervalltyp		KL	EL	KL	EL	Grenzvarianz
$\alpha = 0.01$	$n = 20$	-0.430	-0.430	0.1485	0.1485	
	100	-0.111	-0.116	0.1148	0.1155	0.0099
	1000	-0.032	-0.026	0.0110	0.0107	
$\alpha = 0.05$	$n = 20$	-0.252	-0.283	0.1485	0.1728	
	100	-0.108	-0.086	0.1509	0.1469	0.0475
	1000	-0.039	-0.032	0.0521	0.0503	
$\alpha = 0.1$	$n = 20$	-0.249	-0.216	0.1882	0.1914	
	100	-0.109	-0.083	0.1952	0.1951	0.09
	1000	-0.044	-0.036	0.0902	0.0890	

Wie zu erwarten unterscheiden sich die Werte für die mit $\mathrm{F}_{n,res}$ gebildeten Prognoseintervalle nicht wesentlich von denen für die mit $\mathrm{F}_{n,res,z}$ gebildeten. In beiden Fällen und bei beiden betrachteten Verteilungen nähern sich die Beobachtungen den theoretischen Grenzwerten. Dabei scheint die Abweichung des Mittelwertes im EL-Fall betragsmäßig leicht kleiner zu sein als bei der klassischen Schätzung, und auch die Varianzen stabilisieren sich möglicherweise etwas schneller. Im Fall einer symmetrischen Fehlerverteilung liefert die Konstruktion nichtparametrischer Prognoseintervalle auf Grund des EL-Verteilungsschätzers im übrigen aber auch bei endlichem Stichprobenumfang keine signifikante Verbesserung.

Qualitativ völlig anders sind die Ergebnisse bei der Simulation von oberen Prognoseschranken. Dazu wurden zu obigen Fehlerverteilungen und Parameterwerten je 1000 Schranken der Form

$$\hat{\rho}_n X_n + \mathrm{F}_{n,res}^{-1}(1-\alpha) \tag{KL}$$

und 1000 der Form

$$\hat{\rho}_n X_n + \mathrm{F}_{n,res,z}^{-1}(1-\alpha) \tag{EL}$$

simuliert. Zu jedem Intervall wurden 10000 unabhängige Prognosewerte erzeugt und die mittlere Überschreitungshäufigkeit bestimmt. Da die ρ-Abhängigkeit der Ergebnisse minimal ist, geben wir wieder nur die Ergebnisse für $\rho = 0.0$ in einer Tabelle wieder.

Tabelle 13		$\mathcal{N}(0,1)$		t_3	
Schrankentyp		KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	0.053	0.053	0.052	0.052
	100	0.019	0.014	0.020	0.014
	1000	0.011	0.011	0.011	0.010
$\alpha = 0.05$	$n = 20$	0.107	0.073	0.102	0.074
	100	0.058	0.053	0.060	0.054
	1000	0.051	0.051	0.051	0.050
$\alpha = 0.1$	$n = 20$	0.157	0.122	0.151	0.120
	100	0.108	0.103	0.111	0.104
	1000	0.102	0.101	0.101	0.100

Die Tabelle zeigt unabhängig von der zu Grunde liegenden Fehlerverteilung eine gute Annäherung an das nominelle Überschreitungsniveau. Dabei scheint die Approximation im EL-Fall etwas schneller zu sein. Der Grund für die zu großen Überschreitungshäufigkeiten bei kleinen Stichprobenumfängen wurde bereits oben diskutiert.

Wirklich von Interesse sind aber die 1000 Werte der Größen

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \leq \hat{\rho}_n X_n + \mathrm{F}_{n,res}^{-1}(1-\alpha)|\mathcal{A}_n\right) - (1-\alpha)\right) \tag{KL}$$

und

$$\sqrt{n}\left(\mathrm{P}\left(X_{n+1} \leq \hat{\rho}_n X_n + \mathrm{F}_{n,res,z}^{-1}(1-\alpha)|\mathcal{A}_n\right) - (1-\alpha)\right), \tag{EL}$$

beziehungsweise die aus diesen berechneten Mittelwerte (MW) und Stichprobenvarianzen (SV) als Schätzer für Erwartungswert und Varianz der gemäß Kapitel 5 gegebenen Grenzvariablen. Der Erwartungswert ist, wie man sich leicht überzeugt, bei zentrierter Fehlerverteilung stets gleich Null, während die Grenzvarianz von der Fehlerverteilung F und dem Niveau α abhängt. Sie ist im Fall der Schätzung mittels $\mathrm{F}_{n,res}$ durch

$$\mathrm{f}\left(\mathrm{F}^{-1}(1-\alpha)\right)^2 \sigma^2 + \alpha(1-\alpha) \tag{KL}$$

und im Fall der Schätzung auf Grund von $\mathrm{F}_{n,res,z}$ durch

$$\mathrm{f}\left(\mathrm{F}^{-1}(1-\alpha)\right)^2 \sigma^2 + \alpha(1-\alpha) - \frac{1}{\sigma^2}\mathrm{U}^2\left(\mathrm{F}^{-1}(1-\alpha)\right) \tag{EL}$$

gegeben. Die folgenden Tabellen vergleichen die Simulationsergebnisse untereinander und mit den theoretischen Werten, wobei wir hier wieder nur den Fall $\rho = 0.0$ darstellen. Die zur Berechnung der Grenzvarianzen benötigten Quantile der Standardnormalverteilung und der t_3-Verteilung sind wohlbekannt und zum Beispiel bei Hartung (1991, S.891+892) vertafelt.

Tabelle 14		Fehlerverteilung: $\mathcal{N}(0,1)$					
		MW		SV		Grenzvarianz	
Schrankentyp		KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	-0.192	-0.192	0.0562	0.0562		
	100	-0.094	-0.036	0.0202	0.0125	0.0106	0.0099
	1000	-0.035	-0.018	0.0113	0.0103		
$\alpha = 0.05$	$n = 20$	-0.254	-0.105	0.1148	0.0703		
	100	-0.082	-0.029	0.0646	0.0473	0.0581	0.0475
	1000	-0.038	-0.017	0.0609	0.0465		
$\alpha = 0.1$	$n = 20$	-0.256	-0.097	0.1523	0.1080		
	100	-0.080	-0.033	0.1133	0.0827	0.1208	0.09
	1000	-0.050	-0.024	0.1219	0.0899		

Tabelle 15		Fehlerverteilung: t_3					
		MW		SV		Grenzvarianz	
Schrankentyp		KL	EL	KL	EL	KL	EL
$\alpha = 0.01$	$n = 20$	-0.189	-0.190	0.0737	0.0736		
	100	-0.098	-0.038	0.0215	0.0128	0.0100	0.0084
	1000	-0.030	-0.011	0.0105	0.0083		
$\alpha = 0.05$	$n = 20$	-0.231	-0.106	0.1293	0.0859		
	100	-0.102	-0.039	0.0726	0.0549	0.0537	0.0412
	1000	-0.034	-0.016	0.0505	0.0400		
$\alpha = 0.1$	$n = 20$	-0.227	-0.090	0.1827	0.1265		
	100	-0.108	-0.038	0.1558	0.1223	0.1215	0.0932
	1000	-0.039	-0.019	0.1175	0.0890		

Wir sehen, daß die Simulationsergebnisse bei beiden Verteilungen parallel zu den theoretischen Werten eine Varianzverkleinerung für die EL-basierten Schranken aufweisen. Auch ist der Mittelwert in allen Fällen betragsmäßig kleiner als bei der klassischen Schätzung, was einen weiteren Pluspunkt für die EL-geschätzten Schranken darstellt.

Es zeigt sich also, daß das EL-Prinzip bei Konstruktion von nichtparametrischen Prognoseintervallen bei symmetrischen Verteilungen keine Verbesserung ergibt, bei der Konstruktion von Prognoseschranken aber der klassischen Schätzung klar vorzuziehen ist.

Literaturverzeichnis

1. Babu, G.J./ Rao, C.R.: *Goodness-of-Fit Tests when Parameters are Estimated,* erscheint in Sankhya, Ser. A, 2004

2. Bickel, P.J./ Klaassen, C.A.J./ Ritov, Y./ Wellner, J.A.: *Efficient and Adaptive Estimation for Semiparametric Models,* Johns Hopkins University Press, Baltimore, 1993, Springer, New York, 1998

3. Billingsley, P.: *Convergence of Probability Measures,* Wiley, New York, 1968

4. Boldin, M.V.: *Estimation of the Distribution of Noise in an Autoregression Scheme,* Theor. Probab. Appl. **27**, 866-871, 1982

5. Chuang, C./ Chan, N.H.: *Empirical Likelihood for Autoregressive Models, with Applications to Unstable Time Series,* Statist. Sinica **12**, 387-407, 2002

6. Datta, S./ Sriram, T.N.: *A Modified Bootstrap for Autoregression without Stationarity,* J. Statist. Plann. Inference **59**, 19-30, 1997

7. De la Peña, V.H./ Giné, E.: *Decoupling,* Springer, New York, 1999

8. Devroye, L.: *Non-Uniform Random Variate Generation,* Springer, New York, 1986

9. Durbin, J.: *Weak Convergence of the Sample Distribution Function when Parameters are Estimated,* Ann. Statist. **1**, 279-290, 1973

10. Fink, H.: *Schätzung der Fehlerverteilung in explosiven AR(1)-Prozessen,* Diplomarbeit, Mathematisches Institut der Justus-Liebig-Universität Giessen, 2002

11. Genz M.: *Schätzung der Fehlerverteilung in stabilen autoregressiven Prozessen erster Ordnung,* Diplomarbeit, Mathematisches Institut der Justus-Liebig-Universität Gießen, 2001

12. Ghosh, J.K.: *A New Proof of the Bahadur Representation of Quantiles and an Application,* Ann. Math. Statist. **42**, 1957-1961, 1971

13. Hartung, J.: *Statistik,* 8.Aufl., Oldenbourg-Verlag, München, 1991

14. Koul, H.L./ Levental, S.: *Weak Convergence of the Residual Empirical Process in Explosive Autoregression,* Ann. Statist. **17**, 1784-1794, 1989

15. Koul, H.L.: *Weighted Empiricals and Linear Models,* Institute of Mathematical Statistics, Hayward, 1992

16. Lee, A.J.: *U-Statistics,* Marcel Dekker, New York, 1990

17. Lee, S.: *A Note on the Residual Empirical Process in Autoregressive Models,* Statist. & Probab. Letters **32**, 405-411, 1997

18. Lehmann, E.L.: *Theory of Point Estimation,* Wiley, New York, 1983

19. Ling, S.: *Weak Convergence of the Sequential Empirical Processes of Residuals in Nonstationary Autoregressive Models,* Ann. Statist. **26**, 741-754, 1998

20. Monti, A.C.: *Empirical Likelihood Confidence Regions in Time Series Models,* Biometrika, **84**, 395-405, 1997

21. Müller, U.U./ Schick, A./ Wefelmeyer, W.: *Weighted Residual-Based Density Estimators for Nonlinear Autoregressive Models,* erscheint in Statist. Sinica, 2004

22. Owen, A.: *Empirical Likelihood Ratio Confidence Intervals for a Single Functional,* Biometrika **75**, 237-249, 1988

23. Owen, A.: *Empirical Likelihood Confidence Regions,* Ann. Statist. **18**, 90-120, 1990

24. Owen, A.: *Empirical Likelihood,* Chapman & Hall / CRC, Boca Raton, 2001

25. Press, W.H./ Teukolsky, S.A./ Vetterling, W.T./ Flannery, B.P.: *Numerical Recipes in C,* 2.Ed., Cambridge University Press, 1992

26. Pukelsheim, F.: *Optimal Design of Experiments,* Wiley, New York, 1993

27. Qin, J./ Lawless, J.: *Empirical Likelihood and General Estimating Equations,* Ann. Statist. **22**, 300-325, 1994

28. Shao, J./ Tu, D.: *The Jackknife and Bootstrap,* Springer, New York, 1995

29. Shao, J.: *Mathematical Statistics,* Springer, New York, 1999

30. Stute, W./ Gonzáles-Manteiga, W./ Presedo-Quindimil, M.: *Bootstrap Based Goodness-Of-Fit-Tests,* Metrika **40**, 243-256, 1993

31. Stute, W./ Gründer, B.: *Nonparametric Prediction Intervals for Explosive AR(1)-Processes,* Nonparam. Statist. **2**, 155-167, 1993

32. Van der Vaart, A.W.: *Asymptotic Statistics,* Cambridge University Press, 1998

33. Witting, H./ Nölle, G.: *Angewandte mathematische Statistik,* Teubner, Stuttgart, 1970

34. Ylvisaker, D.: *A Note on the Absence of Tangencies in Gaussian Sample Paths,* Ann. Math. Statist. **39**, 261-262, 1968

35. Zhang, B.: *Estimating a Distribution Function in the Presence of Auxiliary Information,* Metrika **46**, 221-244, 1997

Erklärung

Hiermit erkläre ich, daß ich diese Arbeit selbständig verfaßt und nur die angegebenen Hilfsmittel verwendet habe.

Gießen, den 31.08.2004